



# Diverging Moments and Parameter Estimation

Paulo Gonçalves, Rudolf Riedi

## ► To cite this version:

Paulo Gonçalves, Rudolf Riedi. Diverging Moments and Parameter Estimation. [Research Report] RR-4647, INRIA. 2002. inria-00071938

**HAL Id: inria-00071938**

**<https://inria.hal.science/inria-00071938>**

Submitted on 23 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# *Diverging Moments and Parameter Estimation*

Paulo Gonçalves — Rudolf Riedi

**N° 4647**

Novembre 2002

THÈME 4

 *apport  
de recherche*



## Diverging Moments and Parameter Estimation

Paulo Gonçalves \*, Rudolf Riedi †

Thème 4 — Simulation et optimisation  
de systèmes complexes  
Projet IS2

Rapport de recherche n° 4647 — Novembre 2002 — 38 pages

**Abstract:** Heavy tailed distributions which allow for values far from the mean to occur with considerable probability are of increasing importance in various applications as the arsenal of analytical and numerical tools grows. Examples of interest are the Stable and more generally the Pareto distributions for which moments of sufficiently large order diverge. In fact, the asymptotic powerlaws of the distribution function at infinity and zero are directly related to the existence of positive and negative order moments, respectively. In practice, however, when dealing with finite size data sets of an unknown distribution, standard empirical estimators of moments will typically fail to reflect the theoretical divergence of moments and provide finite estimates for all order moments. The main contribution of this paper is an empirical wavelet-based estimator for the *characteristic exponents* of a random variable, which bound the interval of all orders  $r$  with finite moment.

This objective is achieved by deriving a theoretical equivalence between finiteness of moments and the local Hölder regularity of the characteristic function at the origin and by deriving a wavelet based estimation scheme which is particularly suited to characteristic functions.

**Key-words:** diverging moments, heavy tail distributions, characteristic functions, wavelet transform, multifractal analysis.

\* Inria Rhône-Alpes, projet IS2.

† Dept of ECE, Rice University, Houston (TX), USA.

## Moments divergents et estimation de paramètres

**Résumé :** Les distributions à queues lourdes – caractérisant des variables aléatoires pouvant prendre, avec une probabilité forte, des valeurs très éloignées de la valeur moyenne – ont une importance croissante dans un large éventail d’applications. Les distributions stables, et plus généralement les distributions de Pareto, sont des exemples canoniques pour lesquels les moments d’ordre suffisamment élevé divergent. En fait, les décroissances asymptotiques des distributions en l’infini et en zéro, majorées par des lois de puissance, sont directement liées à l’existence des moments positifs et négatifs des variables aléatoires. En pratique, cependant, lorsque l’on traite des échantillons de taille finie et de distribution inconnue, les estimateurs empiriques de moments, classiquement utilisés, ne décèlent pas cette divergence théorique, et produisent des estimations finies à tous les ordres. L’apport principal de cet article est un estimateur à base d’ondelettes des *exposants caractéristiques* d’une variable aléatoire, bornant l’intervalle sur lequel les moments d’ordre  $r$  existent.

Pour ce faire, nous établissons tout d’abord une équivalence théorique entre la finitude des moments et la régularité hôlderienne de la fonction caractéristique en l’origine; nous proposons ensuite, un schéma d’estimation construit sur la décomposition en ondelettes de la fonction caractéristique.

**Mots-clés :** divergence de moments, distributions leptokurtiques, fonction caractéristique, transformée en ondelettes, analyse multifractale.

# 1 Introduction

Estimators for moments of random variables are abundant. They are not only important for parameter estimation, when the underlying distribution law is known, and for data fitting, but they are also useful at identifying unknown distributions from sample data. For instance the Kurtosis statistic hypothesis test resolves gaussianity versus non gaussianity, whereas for a poisson random variable, mean and variance should be equal, to cite but two. On the other hand, many applications integrate moment estimates as a crucial ingredient. That is the case in multifractal analysis, where the  $q$ -th order absolute moments of the increments (or the wavelet coefficients) of a process, hold valuable information on the local behavior of its paths.

Pathologies emerge when moments are infinite or not defined, like for Cauchy distribution which has infinite second moment and undefined mean. As infinite moments may degrade the performance of estimators (possibly introducing some systematic errors) or reduce the speed of convergence to limiting laws, special attention must be dedicated to their theoretical existence. Now, in multifractal analysis, infinite moments may indicate phase transitions that are highly informative about the process regularity.

All this motivates the development of statistical methods to determine the finiteness of moments of a distribution given finite sample data.

To this end, we propose an approach that combines two facts. First, the characteristic function  $\phi(u) \triangleq \mathbb{E}[\exp(iuX)]$ , being the Fourier transform of the distribution of  $X$ , has as many continuous derivatives at zero as the distribution has finite moments of positive orders. In particular,  $\phi^{(k)}(0) = i^k \mathbb{E}[X^k]$  whenever one of the two is defined [22]. We generalize this fact from integer orders  $k$  to real ones.

Second, the wavelet transform has proved a particularly efficient tool for measuring the local Hölder regularity of a function. In short, if  $\phi$  is Hölder continuous of order  $h$  ( $h > 0$ ) at zero (see text for precise definition) then the same decay holds for the coefficients magnitude of a wavelet with at least  $\lfloor h \rfloor + 1$  vanishing moments. In some sense, the reverse is true as well [5, 13].

Thus, from the decay of the wavelet coefficients of the characteristic function we may estimate the (positive and negative) *critical order*  $\lambda^-$ ,  $\lambda^+$  of a distribution, by which we mean here

$$\begin{aligned} \lambda^+ &\triangleq \sup\{q > 0 : \mathbb{E}[|X|^q] < \infty\} \\ \lambda^- &\triangleq \inf\{q < 0 : \mathbb{E}[|X|^q] < \infty\}. \end{aligned} \tag{1}$$

The recipe of our estimator is dramatically simple:

- from the sample data set  $\{X_n, n = 1, \dots, N\}$  compute the empirical characteristic function  $\widehat{\Phi}(u) = N^{-1} \sum_n \exp\{iuX_n\}$ ;
- compute the wavelet decomposition  $W[\widehat{\Phi}](u, s)$ , and retain solely the coefficient lying at the origin  $\widehat{W}(s) \triangleq W[\widehat{\Phi}](0, s)$ ;
- within some predefined scale intervals, perform a linear regression of  $\log \widehat{W}(s)$  against  $\log s$  to get the two characteristic exponents  $\lambda^-$  and  $\lambda^+$ .

We will demonstrate the effectiveness of this estimator looking at symmetrical stable distributions, which are defined by

$$\phi_X(u) = \mathbb{E}[\exp(iuX)] = \exp(-\sigma^\alpha |u|^\alpha + i\mu u) \quad (2)$$

and which has finite absolute moments of order  $q$  exactly for  $q \in (-1, \alpha)$ . We then continue to apply the estimator in the context of multifractal analysis, where we are able to sharpen the estimates of multifractal spectra of Lévy flight and certain cascades and distinguish between the two processes.

## 2 Background

In this section we collect well known facts on the existence of moments as well as the wavelet analysis of irregular signals.

### 2.1 Estimators for Moments

Most tests for the existence of moments emerge as by-products of parameterized distribution (or characteristic) functions estimators. For instance, Nolan in [21] proposed a maximum likelihood estimator for general alpha-stable laws (including Gaussian and Cauchy) based on a large sample data set. As no closed form exists for these distribution functions (aside from some particular rational values of the parameters), he proposes an efficient numerical resolution of the maximum likelihood equation. Previously, Koutrouvelis [16], Mc Culloch [19, 20], among many others, have proposed two different estimators of the parameters of the same  $\alpha$ -stable laws, either based on Pareto approximation for  $\alpha$ -stables tails, or on the analytic form of the characteristic function (Fourier transform of the distribution). In this particular case of  $\alpha$ -stable distributions, existence of high order moments is determined by the estimation of the index of stability  $\alpha \in (0, 2]$ , as we know that

$$\int |x|^r dF(x) < +\infty, \quad \text{for } r \in (-1, \alpha).$$

Starting from a closed form for the characteristic function (or in some cases a numerical approximation of the density function), all these methods aim at finding the minimum of the log-likelihood function, given the data. As a result, it is well known that those approaches are optimal in the sense of minimum variance, as they yield efficient parameterized estimators attaining the Cramér Rao bound [16, 21, 8].

One major weakness of these parameterized approaches though is when data distribution is unknown. Then, applying parameterized estimators to these data set may lead to misleading estimations whenever the true underlying distributions do not match the model. Here, we propose a *non-parametric* estimator for the decay of fat tail distributions, without any assumption on the density model. Our approach strongly relies on the moment generating function and its local regularity at the origin.

The problem of existing negative moments can be reformulated with a simple change of variables  $x \mapsto x^{-1}$ , as a positive moment existence problem. Then, we could apply our estimator to  $x^{-1}$

instead of  $x$  directly, allowing thus for determining a lower negative bound for the existence of  $\int_{-\infty}^{\infty} x^{-r} dF(x)$ ,  $r > 0$ . However, as we will demonstrate, both positive and negative bounds, can be evaluated at once, using the same procedure applied to the same data set  $\{x_i\}_{i=1, \dots, N}$ .

## 2.2 Characteristic Function and Moments

One common way for calculating high order moments of a random variable  $\mathbf{x}$  is to use the characteristic function:

$$\phi(u) = \mathbb{E}e^{iux} = \int e^{iux} dF(x). \quad (3)$$

Using simple duality argument between time and frequency (via the Fourier transform in (3)), the behavior of this function at the origin relates to the tail behavior of the distribution  $F$  for large  $x$ . In particular, whenever the interger  $p$ th order moment of  $x$  exists, the  $p$ th derivative of  $\phi$  at the origin exists as well and they simply relate as follows

$$\mathcal{M}_p \triangleq \int x^p dF(x) = \frac{d^p}{du^p} \phi(u) \Big|_{u=0} = \phi^{(p)}(0), \quad (4)$$

This justifies  $\phi$  to be also referred to as the *moment generating function*. Vice versa, when  $p$  is even, existence of  $\phi^{(p)}(0)$  implies existence of  $\mathcal{M}_p$ , whereas when  $p$  is odd, existence of  $\phi^{(p)}(0)$  does not necessarily guarantees existence of  $\mathcal{M}_p$  (see for instance example of  $\phi(t) = C^{-1} \sum_{j=2}^{\infty} \cos jt / (j^2 \log j)$  treated in [15, p. 411]).

Clearly, the integrability of  $|x|^p dF(x)$  at infinity, i.e. the rate of the decay of  $F(-x)$  and  $1 - F(x)$  towards zero ( $x \rightarrow \infty$ ), reflects the differentiability of  $\phi$  at the origin.

More precisely, a theorem due to Binmore and Stratton [3, 15] relates the Lipschitz regularity of  $\phi$  at the origin, to the tail decay of the distribution as  $x \rightarrow \infty$

**Theorem 1** *Let  $0 < r < 2$  and let  $\phi(u)$  be the characteristic function of a distribution  $F(x)$ . Then,*

$$1 - \operatorname{Re} \phi(u) = O(u^r) \text{ as } u \rightarrow 0^+ \quad (5)$$

*if and only if*

$$1 - F(x) = O(x^{-r}) \text{ and } F(-x) = O(x^{-r}), \quad x \rightarrow \infty. \quad (6)$$

This theorem tells us that the Lipschitz condition on the characteristic function at the origin is closely related to the algebraic decay of the density function at infinity, from which we can possibly deduce the supremum bound  $\lambda^+$  of existing moments.

To obtain a connection to the moments let us recall the Markov inequality:

$$\mathbb{P}[|X| > a] \leq a^{-r} \mathbb{E}|X|^r, \quad \forall r > 0, \forall a > 0. \quad (7)$$

Consequently, for any  $r > 0$  for which the moment  $\mathcal{M}_r$  (4) is finite, the distribution  $F$  must decay as in (6). Reversely, theorem [15, Th. 11.3.1] implies that whenever  $F$  decays as in(6) with  $r > 0$ , then  $\mathcal{M}_{r'}$  is finite for all  $r' < r$ . To combine these two observations we recall the definition (1) of the (positive) *critical order*  $\lambda^+$  of a distribution and note:



**Lemma 2** *For any  $\rho > 0$ , the next two statements are equivalent:*

(a)  $1 - F(x) = O(x^{-r})$  and  $F(-x) = O(x^{-r})$ ,  $x \rightarrow \infty$ ,  $\forall r < \rho$

(b)  $\mathcal{M}_r = \int_{-\infty}^{\infty} |x|^r dF(x) < \infty$ ,  $\forall r < \rho$

*In other words, the following two expressions for  $\lambda^+$  are equivalent:*

$$\begin{aligned} \lambda^+ &= \sup\{r > 0 : \mathcal{M}_r < \infty\} \\ &= \sup\{r > 0 : 1 - F(x) + F(-x) = O(x^{-r}) \text{ as } x \rightarrow \infty\} \end{aligned} \quad (8)$$

Now, restricting lemma 2 to  $\rho$  in  $(0, 2)$ , and combining it with Theorem 1 we obtain the following corollary:

**Corollary 3** *If either side is known to be strictly less than 2 then the following equation holds:*

$$\lambda^+ = \sup\{r > 0 : 1 - \operatorname{Re} \phi(u) = O(u^r) \text{ as } u \rightarrow 0^+\}. \quad (9)$$

While lemma 2 establishes two entirely equivalent definitions of the positive critical order  $\lambda^+$ , we find from corollary 3 that this important parameter of a distribution coincides with the local Hölder regularity of the characteristic function at the origin, at least when it lies within interval  $(0, 2)$ .

With this in mind, we present wavelet theory in the next section with particular emphasis on their natural abilities to detect and estimate the local regularity of a function. Subsequently, we will use a wavelet approach to obtain certain extensions of corollary 3 for orders higher than 2.

## 2.3 Wavelets and Local Regularity

We introduce Hölder exponents and show how to estimate them using wavelets.

**Local regularity** The concept of local *differentiability* of a function is generalized to non-integer values through the notion of local Hölder (or Lipschitz) regularity as follows [18]

**Definition 4** *A function  $g$  is said to be locally Hölder regular of order  $r$  at time  $t = t_0$ , denoted as  $g \in \mathcal{C}^r(t_0)$ , if there exist a polynomial  $P_r$  of order at most  $\lfloor r \rfloor$  and a constant  $C$  satisfying*

$$|g(t) - P_r(t)| \leq C|t - t_0|^r. \quad (10)$$

We use the index  $r$  for the "Hölder polynomial"  $P_r$  to indicate the order to which it approximates the function  $g$ ; this is typically larger than the order of the polynomial itself.

Clearly,  $P(t_0) = g(t_0)$  and if the Taylor polynomial  $T$  of  $g$  of order  $r$  around  $t_0$  exist then  $P = T$ . In general, though, the Hölder polynomial  $P$  can have higher degree than  $T$  (consider  $g(t) = 1 + t + t^2 + t^{3.1} \sin(1/t)$  with  $T(t) = 1 + t$  but  $P(t) = 1 + t + t^2$ ) and corresponds to the regular part of  $g$ , whereas the singularity behavior is characterized by an algebraic function that bounds the local variations of  $g$  in the vicinity of  $t_0$ .

Similar to the local degree of differentiability of a function  $g$  we define the *local Hölder exponent* or local degree of Hölder regularity  $\mathcal{H}(t_0)$  as:

$$\mathcal{H}_g(t_0) \triangleq \sup\{r > 0 : g \in \mathcal{C}^r(t_0)\} \quad (11)$$

At times we prefer to drop the index and write  $\mathcal{H}(t_0)$  when the choice of function  $g$  is clear from the context.

**Fractal analysis with wavelets** A wavelet analysis consists in a linear decomposition of a signal  $g$  onto a set of analyzing functions<sup>1</sup>

$$\{\psi_{t,s}(u) \triangleq |s|^{-1} \psi((u-t)/s), (t,s) \in \mathbb{R} \times \mathbb{R}_+^*\} \quad (12)$$

through the inner product

$$W[g](t,s) \triangleq \int g(u) \psi_{t,s}(u) du. \quad (13)$$

Conceptually, this transform can be viewed as a partitioning of the time-frequency space, where  $W[g](t,s)$  measures the correlation between  $g$  and each logon  $\psi_{t,s}$ . All of these time-frequency cells  $\psi_{t,s}$  are time-shifted and scale changed versions of a unique prototype function  $\psi$ . Therefore, for the time-frequency tiling to be consistent, the mother wavelet must be localized in the time and in the frequency domain simultaneously. Formally, these constraints transpose to the following: We call a wavelet  $\psi$  *admissible of regularity*  $N_\psi$ , if it has the following three properties:

- $|\psi^{(k)}(t)| \leq C_1(1+|t|)^{-N_\psi-1}$  for  $k = 0, \dots, N_\psi$ ,
- $\int t^k \psi(t) dt$  for  $k = 0, \dots, N_\psi - 1$ , and
- $\int_0^\infty |\Psi(\nu)|^2 / \nu d\nu = \int_0^\infty |\Psi(-\nu)|^2 / \nu d\nu = 1$ .

Now, because equation (13) conveys information on the local oscillatory behavior of the analyzed process  $g$ , it is possible to access the local Hölder exponent  $\mathcal{H}(t)$  from the dynamic of wavelet coefficients across scales, according to [12, 14, 25]:

**Theorem 5** Consider an admissible wavelet  $\psi$  of regularity  $N_\psi$ . Denote by  $W(t,s)$  the wavelet coefficients of a given signal  $g$  with respect to  $\psi$ . The following holds:

- (a) If the signal  $g$  is locally Hölder regular of degree  $r$  at time  $t = t_0$ , i.e.,  $g \in \mathcal{C}^r(t_0)$  and if  $N_\psi \geq \lfloor r \rfloor + 1$ , then there exists a constant  $C$  such that

$$|W(t,s)| \leq C(s^r + |t - t_0|^r) \quad \text{for } s \rightarrow 0^+. \quad (14)$$

In particular, the bound (14) holds for any  $r$  strictly smaller than  $\min(\mathcal{H}_g(t_0), N_\psi)$ .

---

<sup>1</sup>We restrict ourselves to the case of real continuous wavelet transforms, even though all theoretic results we present here transpose directly to the discrete framework of real orthogonal wavelets.

(b) Reciprocally, if there exist numbers  $C$ ,  $r$  and  $r'$  such that  $r' < r < N_\psi$  and such that

$$|W(t, s)| \leq C \left( s^r + \frac{|t - t_0|^{r'}}{\log|t - t_0|} \right), \quad \text{for } s \rightarrow 0^+. \quad (15)$$

then  $g$  is locally Hölder regular of degree  $r$  at time  $t_0$ , in particular  $\mathcal{H}_g(t_0) \geq r$ .

It is noteworthy that the terms  $s^r$  above relate directly to the Hölder regularity of the signal, while the terms  $|t - t_0|$  are connected to the decay of the wavelet. This result is the key to estimating the local Hölder exponent  $\mathcal{H}_g(t_0)$  through the decay of the wavelet coefficients in the vicinity of the singularity. Indeed, up to a correction depending on the distance from the location  $t_0$  of interest, this decay is algebraic in kind with an exponent closely related to  $\mathcal{H}_g(t_0)$ .

In practice, we retain the sole wavelet coefficients lying on the local maxima modulus line that points to the singularity [18] which provide all the information needed.

The issue of estimating the Hölder regularity becomes, thus, particularly simple in the case where the wavelet coefficients  $W(\cdot, a)$  are maximal at  $t_0$  for all scales  $s$ . Indeed, if this is the case, then each of the bounds (14) and (15) holds for all  $t$  if and only if

$$|W(t_0, s)| \leq C s^r \quad \text{for } s \rightarrow 0^+. \quad (16)$$

To arrive at a simple presentation of this particular case we set:

$$\rho^+ := \sup\{r > 0 : |W(t_0, s)| \leq C s^r \text{ for } s \rightarrow 0^+\}. \quad (17)$$

**Corollary 6** Consider a wavelet  $\psi$  of regularity  $N_\psi$  and a signal  $g$  with local Hölder regularity  $\mathcal{H}$  at  $t_0$ . Assume the wavelet coefficients  $W(t, s)$  of the signal  $g$  are maximal at  $t = t_0$  for all scales  $s$ . Provided  $N_\psi > \rho^+$  we have

$$\mathcal{H}_g(t_0) = \rho^+. \quad (18)$$

This result owes its simplicity to the fact that the wavelet coefficients are maximal at  $t_0$ . As a further advantage we note that the relevant condition  $N_\psi > \rho^+$  is entirely determined by the wavelet analysis of the signal and no prior knowledge on the signal is required. This is particularly useful in practical applications.

**Proof**

The claim follows immediately from the above theorem 5. Indeed, (14) holds for any  $r < \mathcal{H}$  showing that any such  $r$  must be smaller than  $\rho^+$ , thus,  $\mathcal{H} \leq \rho^+$  similarly as before. Finally, (15) holds for any  $r < \rho^+$  and  $\rho^+ \leq \mathcal{H}$  similarly as before.  $\diamond$

It remains to indicate whether and when we can expect the relevant condition  $N_\psi > \rho^+$  of corollary 6 to be met. In addition, it is useful to understand the wavelet decay in the case where this condition is not satisfied.

**Lemma 7** Let  $\psi$  have  $N_\psi$  vanishing moments. Consider a signal  $g$  with Hölder regularity  $\mathcal{H}_g(t_0)$ . Then the following two statements hold:

1. If  $N_\psi > \mathcal{H}$  then  $N_\psi > \rho^+$ .
2. Provided that the  $N_\psi$ -th absolute moment  $\int |x|^{N_\psi} |\psi(x)| dx$  is finite, we have:  
If  $N_\psi \leq \mathcal{H}_g(t_0)$  then  $N_\psi \leq \rho^+$ , i.e., there is a constant  $C$  such that

$$|W(t_0, s)| \leq C \cdot |s|^{N_\psi}. \quad (19)$$

We note that wavelet coefficients may decay significantly faster than the order of the wavelet. For examples we refer the interested reader to the study of cascades [27].

**Proof**

Let us assume  $t_0 = 0$  for simplicity.

To establish the first statement we proceed indirectly and assume that  $N_\psi \leq \rho^+$ . Then, (15) holds for any  $r < N_\psi$  and any  $r' < r$ . Letting  $r$  become arbitrarily close to  $N_\psi$  proves that  $\mathcal{H} \geq N_\psi$  and we are done. Note that lemma 7 makes the dual statement: if  $\mathcal{H} \geq N_\psi$  then  $N_\psi \leq \rho^+$ .

Let us turn now to the second statement. By definition of Hölder regularity there exists a polynomial  $P$  of degree  $N_\psi$  such that

$$|g(t) - P(t)| \leq C_1 |t|^{N_\psi}.$$

Let  $Q$  denote the polynomial formed by the terms of order smaller or equal to  $N_\psi - 1$  of  $P$ . Then,  $|g(t) - Q(t)| \leq C_2 |t|^{N_\psi}$  since  $|Q(t) - P(t)|$  is of the order  $|t|^{N_\psi}$ . Since  $Q$  is of degree smaller than  $N_\psi - 1$  we have  $\int Q(t) \psi_{0,s}(t) dt = 0$  and conclude

$$\begin{aligned} |W(0, s)| &= \left| \int (g(t) - Q(t)) \psi_{0,s}(t) dt \right| \leq \int |g(t) - Q(t)| s^{-1} \psi(t/s) dt \\ &\leq \int C_2 |t|^{N_\psi} s^{-1} \psi(t/s) dt = C_2 |s|^{N_\psi} \int |u|^{N_\psi} \psi(u) du. \end{aligned}$$

Since the last integral is finite and does not depend on  $s$  we may absorb it into the constant and the claim is established.  $\diamond$

The first statement (1) says that you can always meet the necessary condition for corollary 6 by choosing a wavelet with sufficient regularity. However, this regularity is typically *not* known in applications. The second statement (2) tells us what to expect when choosing a wavelet with insufficient regularity. This will become important to establish a theoretical result (theorem 16) as well as when discussing the scaling region in numerical applications.

### 3 Estimating the Characteristic Exponent

Analyzing the behavior of the characteristic function around the origin using the wavelet transform we arrive at an extremely simple, yet efficient non-parametric estimator for the decay of tail of a distribution and, thus, the range of finite moments.

### 3.1 Semi-definite wavelet analysis of characteristic functions

Our estimator relies on the strong relation between the Hölder regularity of the real part characteristic function  $\phi$  and the existence of moments.

**Symmetry** To establish this connection, we require a simple result.

**Lemma 8** *For any non-integer  $h$  strictly smaller than  $\mathcal{H}_{\text{Re } \phi}(0)$  the Hölder polynomial  $P_h$  (compare (10)) of the real part of a characteristic function is necessarily even, i.e.,  $P_h(t) = P_h(-t)$ . Similarly, the Taylor polynomial of  $\text{Re } \phi$  has only terms of even order.*

**Proof**

The claim seems only too obvious in view of the fact that  $\text{Re } \phi$  is even as the Fourier transform of a real valued distribution, i.e.,  $\text{Re } \phi(t) = \text{Re } \phi(-t)$ . To provide rigorous arguments, though, recall that the Taylor coefficients are entirely determined by the existing derivatives which must vanish for odd orders. Compare also with (4).

Let us now fix an non-integer  $h < \mathcal{H}_{\text{Re } \phi}(0)$ . By definition,  $\text{Re } \phi$  is then Hölder regular of order  $h$  and we denote the polynomial appearing in (10) by  $P_h$ . We may write

$$\begin{aligned} |P_h(t) - P_h(-t)| &= |P_h(t) - \text{Re } \phi(t) + \text{Re } \phi(-t) - P_h(-t)| \\ &\leq |P_h(t) - \text{Re } \phi(t)| + |\text{Re } \phi(-t) - P_h(-t)| \leq 2C|t|^h. \end{aligned}$$

Being a non-integer,  $h$  is strictly larger than the order of  $P_h(t)$ . Since  $P_h(t) - P_h(-t)$  has an order smaller or equal to the one of  $P_h$  itself, thus, strictly smaller than  $h$ . We conclude that the latter must actually reduce to the zero polynomial, meaning that  $P_h(t) = P_h(-t)$  as claimed.  $\diamond$

Thus motivated we study the wavelet transform of characteristic functions.

**Wavelets with semi-definite Fourier transform** To estimate the Hölder exponent of  $\text{Re } \phi$  at the origin we employ wavelets. To this end, we need to locate the local maxima of the wavelet coefficients  $W[\text{Re } \phi](\cdot, s)$  as a function of location for fixed but arbitrary scale  $s$ .

This turns out to be particularly simple provided the wavelet  $\psi$  has a real-valued semi-definite Fourier transform  $\Psi(\nu) = \int \psi(t) \exp(-it\nu) dt$ . Here, *semi-definite* means that  $\Psi$  does not change sign; in other words,  $\Psi$  is either positive semi-definite, i.e.,  $\Psi(\cdot) \geq 0$ , or it is negative semi-definite, i.e.,  $\Psi(\cdot) \leq 0$ . Examples of such wavelets are the derivatives of even order of the Gaussian kernel: set

$$\psi_p(t) \triangleq c_p \frac{d^{2p}}{dx^{2p}} \exp(-\sigma^2 t^2) \quad (20)$$

where  $c_p$  is a normalization constant and  $p$  is a positive integer. One finds the semi-definite Fourier transform

$$\Psi_p(\nu) = c_p (-1)^p \nu^{2p} \exp\left(\frac{-\pi^2 \nu^2}{\sigma^2}\right). \quad (21)$$

**Lemma 9** *If the Fourier transform  $\Psi$  of the wavelet  $\psi$  is real, square integrable and semi-definite then*

$$|W[\operatorname{Re} \phi](t, s)| \leq |W[\operatorname{Re} \phi](0, s)| \quad (22)$$

*In other words, for fixed scale  $s$  the modulus of the wavelet transform of the real part of a characteristic function is maximal at  $t = 0$  for semi-definite  $\Psi$ .*

**Proof**

Recall Parseval's equality which says that  $\int yZ = \int Yz$  where  $Y$  and  $Z$  are the Fourier transforms of  $y$  and  $z$ . In our application the role of  $y(u)$  is played by  $|s|^{-1}\psi((u-t)/s)$  with Fourier transform  $Y(x) = \Psi(sx) \exp(-itx)$  by simple change of variable. The role of  $Z$  is played by the characteristic function which is the Fourier transform of the distribution  $dF$ . Applying first the definition of  $W$ , then Parseval we get

$$\begin{aligned} W[\operatorname{Re} \phi](t, s) &= \operatorname{Re} \int |s|^{-1} \psi((u-t)/s) \phi(u) du \\ &= \operatorname{Re} \int \Psi(sx) \exp(-itx) dF(x) \end{aligned} \quad (23)$$

Using the simple estimate  $|\operatorname{Re} x| \leq |x|$  as well as the fact that  $\Psi$  is semi-definite and does not change its sign we obtain

$$\begin{aligned} |W[\operatorname{Re} \phi](t, s)| &\leq \int |\Psi(sx) \exp(-itx)| dF(x) = \int |\Psi(sx)| dF(x) \\ &= \left| \int \Psi(sx) dF(x) \right| = |W[\operatorname{Re} \phi](0, s)| \end{aligned} \quad (24)$$

as desired.  $\diamond$

### 3.2 The Basic Form of the Estimator

We may now collect the above results to establish the anticipated tight connection between the existence of moments and the regularity of the real part of the characteristic function as well as a wavelet-based estimator at least for critical orders between 0 and 2. We will address the case of orders larger than 2 subsequently in Section 3.3.

First, combining the form of the Hölder polynomial from lemma 8 with  $\operatorname{Re} \phi(0) = 1$  implies what does not come as a surprise: the Lipschitz continuity condition  $1 - \operatorname{Re} \phi(u) = O(u^r)$  of (5) coincides with the Hölder continuity condition of (10) for  $0 < r < 2$ . Thus, corollary 3 translates to:

**Corollary 10** *If it is known a priori that one of the following expressions is strictly less than 2, then the following equation holds:*

$$\lambda^+ = \mathcal{H}_{\operatorname{Re} \phi}(0). \quad (25)$$

Next, we exploit the advantages of a semi-definite wavelet analysis of characteristic function towards estimating its Hölder exponent at the origin.

**Theorem 11** *Let  $\psi$  be a wavelet with real, semi-definite Fourier transform  $\Psi$  and with regularity  $N_\psi \geq 2$ . If either of the two sides is known to be strictly smaller than 2, then the following equation holds*

$$\lambda^+ = \sup \left\{ r > 0 : \left| \int \Psi(s\nu) dF(\nu) \right| \leq C s^r \quad s \rightarrow 0^+ \right\} \quad (26)$$

**Proof**

Setting  $t = 0$  in (23) we find that the integral expression  $\int \Psi(s\nu) dF(\nu)$  in (26) equals indeed  $W[\text{Re } \phi](0, s)$ . Combining lemma 9 with corollary 6, in particular (18), we conclude that the right hand side of (26) is equal to  $\min(\mathcal{H}_{\text{Re } \phi}(0), N_\psi)$ .

Assume first that  $\lambda^+ < 2$ . Then, corollary 10 says that  $\lambda^+ = \mathcal{H}_{\text{Re } \phi}(0)$ , which must then equal  $\min(\mathcal{H}_{\text{Re } \phi}(0), N_\psi)$  since  $N_\psi \geq 2$ . Thus, (26) is then established.

Second, assume that  $\min(\mathcal{H}_{\text{Re } \phi}(0), N_\psi) < 2$ . Then, it must equal  $\mathcal{H}_{\text{Re } \phi}(0)$  since  $N_\psi \geq 2$ . We conclude again with corollary 10 that  $\lambda^+ = \mathcal{H}_{\text{Re } \phi}(0)$ , and (26) is established also under this assumption.  $\diamond$

Through the above results we are led to *estimate* the critical order of moments  $\lambda^+$  via the exponent  $\rho^+$  of the powerlaw decay of the "semi-definite" wavelet coefficients of  $\text{Re } \phi$  at the origin. Indeed we note:

**Lemma 12** *Provided (26) holds, we have*

$$\lambda^+ = \rho^+ = \limsup_{s \rightarrow 0^+} \frac{\log \int |\Psi(s\nu)| dF(\nu)}{\log(s)} = \limsup_{s \rightarrow 0^+} \frac{\log |W[\text{Re } \phi](0, s)|}{\log s}. \quad (27)$$

**Proof**

First, verify that  $\rho^+$  equals the expression obtained for  $\lambda^+$  in (26). Indeed, since  $\Psi$  does not change sign, we may pull the absolute value into the integral and then use the Parseval formula (24)). Second, let us denote the limsup expressions on the right hand of (27) for the moment by  $\xi$ . Consider  $r < \xi$ ; then the integral decays eventually faster than  $s^r$ , thus,  $r < \rho^+$  and  $\xi \leq \rho^+$ . On the other hand, if  $r > \xi$ , then for any  $C$  the integral is larger than  $C s^r$  for infinitely many  $s > 0$ . Thus,  $r \geq \rho^+$ , and letting  $r$  get arbitrarily close to  $\xi$  we find that  $\xi \geq \rho^+$  and equality is established.  $\diamond$

Theorem 11 states, that this approach should work well since  $\lambda^+ = \rho^+$ , provided  $N_\psi \geq 2 > \rho^+$  and  $\Psi$  is semi-definite.

In a moment, we will devise an estimator of  $\lambda^+$  based on (26). Before doing so, though, we will explore the relation between the critical order and the regularity of  $\phi$  in the general case to obtain an idea, how well this approach will work for  $\lambda^+ > 2$ .

### 3.3 Extending the Estimator to Higher Orders

As outlined in Section 2.1 all three, the asymptotics of the cumulative distribution function  $F$ , the regularity of the characteristic function  $\phi$  and the existence of moments  $\mathcal{M}_r$  are ruled by the same "critical exponent"  $\lambda^+$  as long as this exponent is smaller than 2. Combining this fact with a powerful wavelet analysis we were able to derive the estimator of  $\lambda^+$  in (27).

In the general case, however, only the tail of the distribution  $F$  is a priori connected to the existence of moments, as stated in lemma 2. To connect a critical order  $\lambda^+$  larger than 2 also to the *regularity* of the characteristic function  $\phi$  we need to proceed in several steps. We begin by generalizing a theorem given by Kawata (see [15]) which provides a necessary and sufficient condition for the existence of moments in terms of some kind of averaged regularity of  $\phi$  around the origin.

**Theorem 13** *Let  $p$  be a positive integer and let  $\beta > 0$  such that  $2p < \beta < 2p + 2$ . Assume that  $\phi^{(2p)}(0)$  exists, or equivalently  $\int x^{2k} dF(x) < \infty$  for all integer  $k \leq p$ . Then*

$$\int_{-\infty}^{\infty} |x|^\beta dF(x) < \infty, \quad (28)$$

*if and only if*<sup>2</sup>

$$\int_0^\infty \frac{1}{t^{\beta+1}} |Q_{2p}(t)| dt < \infty, \quad (29)$$

*where*

$$Q_{2p}(t) = \operatorname{Re} \phi(t) - 1 - \sum_{k=1}^p \frac{t^{2k}}{(2k)!} \phi^{(2k)}(0) \quad (30)$$

*is the Taylor rest-term of the characteristic function at zero.*

This theorem is instrumental in relating the existence of higher order moments to the local regularity of the characteristic function. Yet, this result provides not as sharp a connection as theorem 1 does for moments of order smaller than 2. Indeed, a bound on the Taylor rest-term of the form  $|Q_{2p}(t)| \leq C \cdot |t|^{\beta+\varepsilon}$  ( $t \rightarrow 0$ ) implies the convergence of the integral (29) but not necessarily vice versa.

An added difficulty arises from the fact that a wavelet analysis allows to estimate the Hölder regularity via corollary 6, but not the degree of differentiability. While we know of no remedy for the imprecision induced by (29) we are able to overcome the difficulty with derivatives.

In summary, the difficulties to estimate the characteristic exponent  $\lambda^+$  using a wavelet analysis of the characteristic function are twofold. First, to bound  $\lambda^+$  from above we will assume the finiteness of certain moments which immediately implies the existence of derivatives of  $\phi$ . The above theorem will, however, allow only to conclude on some averaged regularity of  $\phi$  in the sense of the convergence of the integral (29), but it will not guarantee a minimal degree of Hölder regularity.

<sup>2</sup>Note that  $|\operatorname{Re} \phi(t)| \leq |\phi(t)| \leq \phi(0) = 1$ , whence  $|Q_{2p}(t)| \leq C \cdot |t|^{2p}$  ( $t \rightarrow \infty$ ). Thus, the convergence of the integral depends solely on the behavior of the integrand around zero.



Second, to bound  $\lambda^+$  from below we must show that certain moments exist. To this end, we must not only demonstrate a certain decay of the Taylor rest-term  $Q_{2p}$  to guarantee (29) but first of all the existence of the derivative of order  $2p$  of the characteristic function. Notably, the existence of derivatives does not follow automatically from a wavelet analysis. To overcome this general difficulty we will extend theorem 11 to provide a means to guarantee the existence of derivatives of the characteristic function from the decay of its wavelet coefficients. We should emphasize that our approach relies heavily on the special framework we are about to present and will not succeed in general. Indeed, counterexamples of functions with Hölder regularity larger than 1, yet without derivative are easily constructed using the idea of chirps.

Before providing this extension, let us establish the above theorem 13.

**Proof**

We follow closely the proof of [15, Thm. 11.4.4.]. Set

$$L_p(x) = \cos(x) - 1 - \sum_{k=1}^p (-1)^k \frac{x^{2k}}{(2k)!} \quad (31)$$

for convenience. If the absolute moment of order  $2p$  exist, or equivalently, if the derivative of order  $2p$  of  $\phi$  at zero exists, then  $L_p$  is integrable with respect to measure  $dF(x)$  and one finds, plugging simply the definition of  $\phi$  as well as the well-known relation between moments and derivatives of the characteristic function (4) into the expression (30) for the Taylor rest-term of  $\phi$ :

$$\begin{aligned} Q_{2p}(t) &= \int_{-\infty}^{\infty} \cos(xt) dF(x) - 1 - \sum_{k=1}^p (-1)^k \frac{t^{2k}}{(2k)!} \int_{-\infty}^{\infty} x^{2k} dF(x) \\ &= \int_{-\infty}^{\infty} L_p(xt) dF(x). \end{aligned} \quad (32)$$

Since  $2p < \beta < 2p + 2$  the function  $L_p(t)/t^{\beta+1}$  is Lebesgue integrable over  $\mathbb{R}$  and some manipulation (see [15, p. 431]) yields

$$\int_0^{\infty} L_p(xt)/t^{\beta+1} dt = |x|^{\beta} \cdot \int_0^{\infty} L_p(t)/t^{\beta+1} dt. \quad (33)$$

This equality is the key. Assume first (29), i.e., assume that  $Q_{2p}(t)/t^{\beta+1}$  is Lebesgue integrable. Integrating (32) with respect to  $t$ , Fubini's theorem allows to exchange the order of integration:

$$\begin{aligned} \int_0^{\infty} Q_{2p}(t)/t^{\beta+1} dt &= \int_0^{\infty} \int_{-\infty}^{\infty} L_p(xt) dF(x) 1/t^{\beta+1} dt \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} L_p(xt)/t^{\beta+1} dt dF(x). \end{aligned}$$

Moreover, Fubini implies that the inner integral as a function of  $x$  is integrable with respect to  $dF(x)$ . Applying (33) we conclude that the inner integral reduces to  $|x|^{\beta}$  times a constant. Thus, we obtain

(28) and the absolute moment  $\mathcal{M}_\beta$  of order  $\beta$  exists. Solving for  $\mathcal{M}_\beta$  we find the explicit formula:

$$\mathcal{M}_\beta = \frac{\int_0^\infty Q_{2p}(t)/t^{\beta+1} dt}{\int_0^\infty L_p(t)/t^{\beta+1} dt}. \quad (34)$$

Vice versa, assume now that (28) holds, i.e., assume that the absolute moment of order  $\beta$  exists. This means that we may integrate (33) with respect to  $dF(x)$ . Similarly as before, Fubini's theorem allows to change the order of integration and implies that  $Q_{2p}(t)/t^{\beta+1}$  is Lebesgue integrable, i.e., (29) holds.  $\diamond$

We formulate the anticipated extension of theorem 11 for a very particular class of wavelets. This is for convenience only and the result could be extended.

**Lemma 14** *Denote by  $\psi_\eta$  the wavelet with Fourier transform of the form  $\Psi_\eta(\nu) = c_\eta |\nu|^\eta \exp(-\pi^2 \nu^2 / \sigma^2)$  for some  $\eta > 0$  and some  $c_\eta > 0$ . Then,*

$$\sup_{s \rightarrow 0} \frac{W[\text{Re } \phi](0, s)}{s^\eta} = c_\eta \int |x|^\eta dF(x) \quad (35)$$

whenever either of the two sides is finite.

Before we provide a proof three remarks are in order.

**Remark 15 (Using fractal wavelets)**

1. The left hand term in (35) is actually a limit, since the sequence is increasing.
2. As an advantage (35) carves a clearly sharper asymptotic behavior than the condition used in the definition of  $\rho^+$  (compare (27) and (26)).
3. To some disadvantage, computing the wavelet transform of  $\psi_\eta$  for one given  $\eta$ , this lemma 14 allows only to conclude whether  $\lambda^+ \geq \eta$  or  $\lambda^+ \leq \eta$ , depending on whether the expressions in (35) are finite or not.



We should note further that such "fractal wavelets" as  $\psi_\eta$  are less practical in general since an explicit form of  $\psi_\eta(t)$  is hard to come by for non-integer  $\eta$ . For our purposes it will serve well, though, since we work in Fourier space entirely.

**Proof**

To begin with, recall (23) which states for  $t = 0$ :  $W[\text{Re } \phi](0, s) = \int \Psi(s\nu) dF(\nu)$ . Next, note that

$$\frac{\Psi(\nu s)}{(s^\eta c_\eta)} = |\nu|^\eta \exp(-(\pi \nu s)^2 / \sigma^2) \nearrow |\nu|^\eta \quad (s \rightarrow 0) \quad (36)$$

for all  $\nu$ . Now, the claim follows immediately from the monotonous convergence theorem of increasing sequences of positive integrable functions (sometimes referred to as the Theorem of Beppo-Levy).

A much more general version of this lemma could be established using arguments along the lines of Thm. 11.2.1 [15].  $\diamond$

Instrumented with lemma 14 we may now summarize our approach to dealing with higher order moments with the following results. Recall  $\rho^+$  from (27).

**Theorem 16** *Let  $\psi$  be any wavelet with  $N_\psi$  vanishing moments and semi-definite, real Fourier transform. If  $N_\psi > \rho^+$  then*

$$\lambda^+ \geq \rho^+ = \mathcal{H}_{\text{Re } \phi}(0). \quad (37)$$

*In words, the critical order of a distribution is bounded from below by the exponent of the powerlaw decay of the wavelet coefficients of the characteristic function at zero.*

As is stated in corollary 6 the condition  $N_\psi > \rho^+$  holds certainly if  $N_\psi > \mathcal{H}_{\text{Re } \phi}(0)$ .

**Proof**

Fix a non-integer  $\beta < \rho^+$ ; note that we may choose  $\beta$  arbitrarily close to  $\rho^+$ , even if  $\rho^+$  should be integer. We will use theorem 13 to show that the moment of order  $\beta$  is finite. This will provide the desired lower bound to  $\lambda^+$ . To this end, we are required to show that the Taylor polynomial exists and that the rest-term has the appropriate decay to make the relevant integral (29) converge (see theorem 13).

Let us start by establishing the regularity of  $\text{Re } \phi$ . Applying the definition of  $\rho^+$  we deduce that  $|W[\text{Re } \phi](0, s)| = O(|s|^\beta)$  (compare (27)). Due to lemma 9 we may apply corollary 6 to  $\text{Re } \phi$  and conclude together with  $N_\psi > \rho^+$  that  $\mathcal{H}_{\text{Re } \phi}(0) = \rho^+$ . Since  $\beta < \rho^+$ , there exists the Hölder polynomial  $P_\beta$  of order at most  $\lfloor \beta \rfloor$  such that

$$|\text{Re } \phi(t) - P_\beta(t)| \leq C|t|^\beta. \quad (38)$$

Next, we would like to establish that  $P_\beta$  is actually the correct Taylor polynomial; and (38) will then provide the necessary bound for the rest-term. We do so in three steps.

First,  $P_\beta$  contains only terms of even power due to lemma 8.

Second, let  $2p$  be the largest even integer smaller than  $\beta$ . We will apply lemma 14 with  $\eta = 2p$  in order to show that the moment of order  $2p$  is finite. Denote the wavelet of lemma 14 by  $\psi_{2p}$ . Since  $\eta = 2p$  is even here, the wavelet  $\psi_{2p}$  coincides with the  $\eta$ -th derivative of  $\exp(-\sigma^2 t^2)$  introduced in (20).

Clearly,  $\psi_{2p}$  has  $2p$  vanishing moments:  $\int t^k \psi_{2p}(t) dt = 0$  for  $k = 0, \dots, 2p-1$  which is easily established using integration by parts  $2p$  times. Due to the exponential decay of  $\psi_{2p}(t)$  for large  $|t|$  and the regularity of  $\text{Re } \phi$  (38) the assumptions of lemma 7 are met and we conclude that

$$\sup_s \frac{W_{2p}(0, s)}{s^{2p}} < \infty$$

Lemma 14 implies the existence of the moment of order  $2p$ .

Third, this fact tells us through (4) that the derivative of order  $2p$  of  $\phi$  exists. Consequently, the Taylor polynomial  $T(t)$  of order  $2p$  exists. It must coincide with the Hölder polynomial  $P_\beta$  since, due to lemma 8, the latter can have no odd terms<sup>3</sup>.

Having accomplished this, we have actually shown that the polynomial  $Q_{2p}$  of theorem 13 exists and that

$$|Q_{2p}(t)| = |\operatorname{Re} \phi(t) - P_\beta(t)| \leq C|t|^\beta.$$

Consequently, the integral  $\int_0^1 Q_{2p}/t^{1+\beta'} dt$  converges for all  $\beta' < \beta$ . Since  $|\operatorname{Re} \phi(t)| \leq 1$  for all  $t$ , we may bound  $Q_{2p}$  by a polynomial of degree  $2p$ . Provided that  $2p < \beta'$  the integral  $\int_1^\infty Q_{2p}/t^{1+\beta'} dt$  converges as well. Due to theorem 13, the moments of order  $\beta'$  exist whenever  $2p < \beta' < \beta$ . Letting  $\beta' \rightarrow \beta$  and then  $\beta \rightarrow \rho^+$  proves that  $\lambda^+$  must be larger than  $\rho^+$ .  $\diamond$

**Corollary 17** *Let  $\mathcal{H}$  denote the Hölder exponent of  $\operatorname{Re} \phi$  at zero. Then,*

$$\mathcal{H} \leq \lambda^+ \leq \lfloor \mathcal{H} \rfloor + 1. \quad (39)$$

Note that  $\lfloor \mathcal{H} \rfloor + 1$  is the smallest integer strictly larger than  $\mathcal{H}$ .

**Proof**

Using a wavelet of regularity larger than  $\mathcal{H}$ , theorem 16 together with theorem 5 gives  $\lambda^+ \geq \mathcal{H}$ . On the other hand, assume that the moment of order  $\lfloor \mathcal{H} \rfloor + 1$  was finite. Then,  $\phi$  would be differentiable  $\lfloor \mathcal{H} \rfloor + 1$  times and would be at least of Hölder regularity of this degree, in contradiction to  $\lfloor \mathcal{H} \rfloor + 1 > \mathcal{H}$ .  $\diamond$

Let us summarize what we learned about our approach which is to estimate the critical order  $\lambda^+$  through the decay rate  $\rho^+$  of wavelet coefficients:

- The wavelet  $\psi$  should possess a real, semi-definite Fourier transform  $\Psi$ ;
- the wavelet  $\psi$  should be of sufficient order; it is of advantage, therefore, to perform the analysis with various wavelets of different regularity;
- in the case  $\lambda^+ < 2$  the method is precise in the sense that  $\lambda^+ = \rho^+$ ;
- in general, our estimator  $\rho^+$  provides a lower bound of  $\lambda^+$  and can guarantee the existence of moments; in addition,  $\lambda^+ \leq \lfloor \rho^+ \rfloor + 1$ ;
- using the "fractal wavelets"  $\psi_\eta$  the critical order  $\lambda^+$  can be estimated as precisely as desired, yet at the cost of computational effort (see remark 15).

---

<sup>3</sup>It is well known that  $\operatorname{Re} \phi(t) - T(t) = o(|t|^{2p})$  for small  $|t|$ . We conclude that

$$|T(t) - P_\beta(t)| \leq |T(t) - \operatorname{Re} \phi(t)| + |P_\beta(t) - \operatorname{Re} \phi(t)| = o(|t|^{2p}) + O(|t|^\beta),$$

and  $P_\beta$  can, thus, differ from  $T$  at best by a term of order  $t^{2p+1}$ . However,  $P_\beta$  contains no odd terms and  $P_\beta = T$ .

### 3.4 Extension of the Estimator for Negative Orders

We are now interested in estimating the negative critical order  $\lambda^-$  defined in (1), of a random variable  $X$  with density  $dF_X(x)$ . Let us define a new random variable  $Y$  using the one to one mapping from  $\mathbb{R}$  to  $\mathbb{R}$ :  $Y = g(X) = X^{-1}$ . Fixing  $Y = y$ , equation  $y = g(x)$  has only one root  $x = y^{-1}$ , and  $|g'(x)| = y^2$ , from which we deduce the distribution of  $Y$ ,  $dF_Y(y) = y^{-2}dF_X(y^{-1})$ . The negative  $q$ th order of  $X$

$$\begin{aligned} \mathcal{M}_q &= \int_{-\infty}^{\infty} x^q dF_X(x), \quad q < 0, \\ &= \int_{-\infty}^{\infty} (y^{-1})^q y^{-2} dF_X(y^{-1}) \\ &= \int_{-\infty}^{\infty} y^{-q} dF_Y(y), \quad -q > 0, \end{aligned} \tag{40}$$

simply corresponds to the positive  $-q$ th order of random variable  $Y$ . Therefore, to estimate  $\lambda^-(X)$  of  $X$ , we can directly apply general results obtained in Section 3.3 for positive higher orders, to get  $\lambda^+(Y) = -\lambda^-(X)$ .

## 4 Practical considerations

In this section, we elaborate on the implementation of our estimator for  $\lambda^+$ , in particular the choice of wavelet and scales to consider, its bias, robustness and optimality properties.

### 4.1 Implementation

Given  $X_i$ , ( $i = 1, \dots, N$ ) a set of  $N$  observed i.i.d. samples of the distribution  $dF(x)$ , we use the empirical estimator for the characteristic function

$$\hat{\phi}(u) \triangleq \phi_N(u) \triangleq \frac{1}{N} \sum_{k=1}^N \exp\{iuX_k\} \tag{41}$$

For our purpose, we need to evaluate this function on a properly sampled interval  $u_j = j \cdot \delta u$ ,  $j = 0, \dots, K-1$ , that we will make more precise later.

We now recall some convergence properties of this empirical characteristic function (see [9, 10] for details), justifying its use in the rest of our method. First,  $\phi_N(u)$  converges almost surely when  $N$  goes to infinity towards  $\phi(u)$  in the  $L^\infty$  sense, over some finite interval  $T$

$$\sup_{|u| \leq T} |\phi_N(u) - \phi(u)| \rightarrow 0. \tag{42}$$

Second, consider the random process  $Y_N(u) = \sqrt{N}(\phi_N(u) - \phi(u))$  and let  $Y(u) = \overline{Y}(-u)$  be a zero mean complex Gaussian process with covariance structure  $\mathbb{E}Y(u)Y(v) = \phi(u+v) - \phi(u)\phi(v)$ . Then, the sequence  $Y_N(u_1), Y_N(u_2), \dots, Y_N(u_m)$  converges in distribution to  $Y(u_1), Y(u_2), \dots, Y(u_m)$ .

It is also shown in [10] that  $Y_N(u)$  converges weakly towards  $Y(u)$  in any finite interval, provided that  $\mathbb{E}|X|^{1+\delta} < \infty$ .

Next to consider is the wavelet decomposition of  $\phi_N(u)$  which simplifies to

$$\begin{aligned}
 W[\phi_N](t, s) &= \int \psi_{t,s}(v) \phi_N(v) dv \\
 &= \frac{1}{N} \sum_k \int \psi_{t,s}(v) \exp\{iX_k v\} dv \\
 &= \frac{1}{N} \sum_k \int |s|^{-1} \psi\left(\frac{v-t}{s}\right) \exp\{iX_k v\} dv \\
 &= \frac{1}{N} \sum_k \int \psi(u) \exp\{iX_k(su + t)\} du \\
 &= \frac{1}{N} \sum_k \exp\{iX_k t\} \int \psi(u) \exp\{iX_k su\} du \\
 &= \frac{1}{N} \sum_k \Psi(s \cdot X_k) \exp\{iX_k t\}.
 \end{aligned}$$

Assuming that  $\Psi$  is real, semi-definite we finally arrive at the surprisingly simple estimator for the maximal wavelet coefficient of  $\text{Re } \phi$  of scale  $s$ , which is the main ingredient in our method:

$$\widehat{W}(s) \triangleq W[\text{Re } \phi_N](0, s) = (1/N) \sum_{k=1}^N \Psi(s \cdot X_k). \quad (43)$$

Finally, according to theorem 16, the characteristic exponent  $\lambda^+$  is estimated from the powerlaw exponent  $\rho^+$  which steers the decay of  $\widehat{W}(s)$ . Taking the logarithmic of this powerlaw model yields the linear trend

$$\log \widehat{W}(s) \approx \widehat{\rho}^+ \log s + \log C, \quad (44)$$

where  $\widehat{\rho}^+$  is simply obtained via a standard (weighted) linear regression procedure of  $\log \widehat{W}(s)$  against  $s$  restricted to some scaling interval  $(s_{\min}, s_{\max})$  to be specified.

Since we assume nothing on the distribution we obtain thus a non-parametric estimator.

## 4.2 One point Statistics

Let us study the bias of both, the simple estimator of the wavelet coefficient (43) itself, as well as the derived estimation of the scaling exponent (44).

Since all observations are drawn from the same distribution, we may write

$$\mathbb{E}[\widehat{W}(s)] = \frac{1}{N} \sum_{k=1}^N \mathbb{E}[\Psi(sX_k)] = \mathbb{E}[\Psi(s \cdot X)] = \int \Psi(sx) dF(x). \quad (45)$$

This shows that as an estimator of the wavelet coefficient  $W(0, s)$  itself,  $\widehat{W}(s)$  is unbiased. However, as we will show, a bias is introduced as we estimate the powerlaw decay of  $W(0, s)$  through the powerlaw decay of  $\widehat{W}(s)$ . This result is similar to the one we got in [1] where we showed that using log-periodograms (Welch estimator) to analyze processes with spectra of the type  $\Gamma_X(f) \sim \sigma^2 |f|^{-\alpha}$  leads to a systematic bias on the estimate of  $\alpha$ . On the other hand, using a wavelet-based spectral analysis (the frequency marginal of a wavelet decomposition) yields an asymptotically unbiased estimator for exponent  $\alpha$ . This is due to the constant relative bandwidth of wavelets that performs a logarithmic tiling of the time axis. The resulting *time-band* analysis has joint time and frequency resolutions that match naturally powerlaw decays as in  $\Gamma_X(f)$ , or in our case, in  $\phi(u)$  around the origin.

**Estimating the critical order: A showcase** To explore the properties of an estimator of the exponent  $\rho^+$  of the wavelet coefficients we first treat a simple case where we assume that

- the distribution is Pareto, i.e.,  $F'(x) = p_X(x) = c_0 x^{-\alpha-1}$  for  $|x| > \delta$  and vanishes elsewhere, with  $c_0 = \alpha \delta^\alpha$ ;
- the wavelet is bandlimited, actually require that  $\Psi(\nu) = 0$  for  $|\nu| \leq \underline{\nu}$ , where  $0 < \underline{\nu}$  is some constant.

Such wavelets are known to exist.

Inserting the particular form of  $p_X(x/s)$  into the bias formula (45) we can extract the scale  $s$  through a substitution. Provided  $s$  is *small enough*, i.e.,  $s < \underline{\nu}/\delta$ , the remaining integral is independent of the scale due to the band limitation of  $\Psi$ . This reads as:

$$\begin{aligned} \mathbb{E}[\widehat{W}(s)] &= \int_{\delta}^{\infty} \Psi(xs) c_0 x^{-\alpha-1} dx = s^\alpha \cdot \int_{\delta s}^{\infty} \Psi(y) c_0 y^{-\alpha-1} dy \\ &= s^\alpha \cdot \int_{\underline{\nu}}^{\infty} \Psi(y) c_0 y^{-\alpha-1} dy = C_\Psi(\alpha) \cdot s^\alpha. \end{aligned} \quad (46)$$

Thus, the exact powerlaw of the density translates into one of  $\widehat{W}(s)$  thanks to the band limitation of the wavelet. Apart from this show-case, approximatively the same decay of  $\widehat{W}(s)$  can be observed under much less restrictive assumptions, as we are about to show.

**Estimating the critical order of fat tail distributions** We relax the above assumptions to the following scenario:

- We consider a simple, heavy tailed probability density function which is symmetrical, constant around the origin and which follows an exact powerlaw in the tails:

$$F'(x) = p_X(x) = \begin{cases} c_1 & \text{if } |x| < \delta, \\ c_2 |x|^{-\alpha-1} & \text{if } |x| \geq \delta, \end{cases} \quad (47)$$

where

$$c_1 = \frac{1}{2\delta} \cdot \frac{\alpha}{\alpha + v} \quad \text{and} \quad c_2 = \frac{\delta^\alpha}{2} \cdot \frac{\alpha}{1 + \alpha/v}.$$

- The wavelet is sufficiently regular:

$$\Psi(\nu) \leq d_\psi |\nu|^{N_\psi}. \quad (48)$$

Let us comment on this choice. Despite its special form, this distribution will be sufficient to explore general fat tail distributions. Clearly, it has finite moments of the orders between  $\lambda^- = -1$  and  $\lambda^+ = \alpha$ . Also,  $v = p_X(\delta)/p_X(0)$  is the ratio of the tail amplitude to the constant value around zero. Clearly, the bound (48) is most efficient for small  $\nu$ , as  $\Psi$  is integrable and must decay at infinity.

To show that  $\mathbb{E}[\widehat{W}(s)]$  scales as  $s^\alpha$ , we need to generalize (46) and split the integral of (45) into two parts,  $|x| < \delta$  and  $|x| \geq \delta$ . We claim that the first part is of the order  $s^{N_\psi}$  while the second term behaves as  $s^\alpha$  plus a term of the order  $s^{N_\psi}$ . In summary, the wavelet estimator decays indeed as  $s^\alpha$  with an error term in the order  $s^{N_\psi}$ , which may introduce a bias in the estimation of  $\lambda^+ = \alpha$ .

Applying (48) we find

$$\int_{-\delta}^{\delta} \Psi(sx) p_X(x) dx \leq c_1 \cdot \int_{-\delta}^{\delta} \Psi(x) dx \leq c_1 \cdot 2\delta \cdot d_\psi \delta^{N_\psi} \cdot s^{N_\psi} \quad (49)$$

as claimed. Next, similarly to (46) we obtain

$$\int_{\delta}^{\infty} \Psi(sx) p_X(x) dx = \frac{1}{s} \int_{s\delta}^{\infty} \Psi(y) p_X(y/s) dy = s^\alpha \cdot c_2 \int_{s\delta}^{\infty} \Psi(y) y^{-\alpha-1} dy. \quad (50)$$

To the contrary of (46) this integral depends on  $s$ . Thus, we write it as  $\int_0^\infty - \int_0^{s\delta}$ . The first term is now a constant, leading to the announced behavior as  $s^\alpha$ . To estimate the second term, we estimate  $\Psi$  in a way similar to (49):

$$\int_0^{s\delta} \Psi(y) y^{-\alpha-1} dy \leq d_\psi \int_0^{s\delta} y^{N_\psi} y^{-\alpha-1} dy = d_\psi \frac{(s\delta)^{N_\psi-\alpha}}{N_\psi-\alpha}. \quad (51)$$

Collecting (49) - (51) we find that  $\mathbb{E}[\widehat{W}(s)] = As^\alpha + O(s^{N_\psi})$ . Bounding the error relied on the regularity (48) of the wavelet, while the exact scaling derives directly from the exact powerlaw (47) of the distribution. We generalize this result as follows:

**Proposition 18** Assume that  $\Psi$  is positive semi-definite. Assume, the distribution has a density  $p_X$  which can be bounded as follows:

$$p_X(x) \begin{cases} \leq a|x|^{-\alpha-1} & \text{for } |x| \geq \delta, \\ \geq b|x|^{-\alpha-1} & \text{for } |x| \geq \delta, \\ \leq c & \text{for } |x| \leq \delta. \end{cases} \quad (52)$$

Assume that the regularity of the wavelet  $\psi$  is larger than the critical order, i.e.,  $N_\psi \geq \alpha$ . Then,

$$\tilde{a} \cdot s^\alpha + O(s^{N_\psi}) \geq \mathbb{E}[\widehat{W}(s)] \geq \tilde{b} \cdot s^\alpha + O(s^{N_\psi}), \quad (53)$$

with  $\tilde{a}/\tilde{b} = a/b$ .



**Proof**

Since  $\psi$  has  $N_\psi$  vanishing moments we know that (48) holds. Proceeding as before, we write

$$\mathbb{E}[\widehat{W}(s)] = \int_{-\delta}^{\delta} \Psi(sx)p_X(x)dx + s^{-1} \cdot \int_{|y|>s\delta} \Psi(y)p_X(y/s)dy$$

The first term is bounded from above as in (49), while the second term maybe framed using the tail bounds on  $p_X$  as

$$s^\alpha \cdot aI \geq s^{-1} \cdot \int_{|y|>s\delta} \Psi(y)p_X(y/s)dy \geq s^\alpha \cdot bI,$$

where

$$I = \int_{|y|>s\delta} \Psi(y)|y|^{-\alpha-1}dy = \int_{-\infty}^{\infty} \Psi(y)|y|^{-\alpha-1}dy - \int_{-s\delta}^{s\delta} \Psi(y)|y|^{-\alpha-1}dy.$$

Here, the last term can be bounded as in (51). In summary:

$$\mathbb{E}[\widehat{W}(s)] = As^\alpha + Bs^{N_\psi} \quad (54)$$

where

$$a \cdot \int_{-\infty}^{\infty} \Psi(y)|y|^{-\alpha-1}dy \geq A \geq b \cdot \int_{-\infty}^{\infty} \Psi(y)|y|^{-\alpha-1}dy \quad (55)$$

which could be estimated further using (48), and where

$$B \leq 2\delta^{N_\psi+1} \cdot c \cdot d_\psi + 2ad_\psi \frac{\delta^{N_\psi-\alpha}}{(N_\psi-\alpha)}. \quad (56)$$

◇

### 4.3 Numerical robustness

Provided that the observations  $X_k$  ( $k = 1, \dots, N$ ) are un-correlated one finds easily

$$\text{var } \widehat{W}(s) = (1/N) \text{var } (\Psi(sX)). \quad (57)$$

Moreover, under the assumptions of proposition 18 we conclude that  $\mathbb{E}\Psi(sX) \sim s^\alpha$  and, considering  $\Psi^2$  as a wavelet,  $\mathbb{E}\Psi^2(sX) \sim s^\alpha$ . Thus,

$$\begin{aligned} \text{var } (\Psi(sX)) &= \mathbb{E}\Psi^2(sX) - (\mathbb{E}\Psi(sX))^2 \\ &\sim s^\alpha \mathbb{E}\Psi^2(X) - s^{2\alpha} (\mathbb{E}\Psi(X))^2. \end{aligned}$$

To provide a more rigorous error estimate let us assume that

$$\Psi(x) = \begin{cases} A_\psi, & x \in I_\psi = [\nu_\psi - \frac{1}{2}N_\psi^{-1/2}, \nu_\psi + \frac{1}{2}N_\psi^{-1/2}] \\ 0, & \text{otherwise.} \end{cases} \quad (58)$$

This somewhat crude boxcar approximation for the wavelet becomes reasonably accurate for the derivatives of the Gaussian kernel  $\psi_p$  (20) as we set  $\nu_\psi = \sqrt{p}\sigma/\pi$ . Indeed,  $\psi_p$  reaches its maximal value  $c_p(p\sigma^2/\pi^2)^p \exp(-p)$  at this  $\nu_\psi$ . Clearly, the approximation becomes more accurate as the regularity increases, i.e.,  $N_\psi \rightarrow \infty$ .

For the box-car wavelet we get

$$\text{var}\widehat{W}(s) = (1/N)\text{var}(\Psi(sX)) = \frac{A_\psi^2}{N} (p_X[sX \in I_\psi] - p_X^2[sX \in I_\psi]).$$

Assuming an exact powerlaw for the tail as in (47) we may write, provided the scale is sufficiently small, i.e.,  $s < (\nu_\psi - \sqrt{N_\psi}/2)/\delta$ :

$$p_X[sX \in I_\psi] = \int_{(\nu_\psi - \frac{1}{2}N_\psi^{-1/2})/s}^{(\nu_\psi + \frac{1}{2}N_\psi^{-1/2})/s} c_2 x^{-\alpha-1} dx = s^\alpha \cdot c_2 \int_{I_\psi} y^{-\alpha-1} dy.$$

Using the mean value theorem we may rewrite the integral by  $y_\psi^{-\alpha-1} \cdot N_\psi^{-1/2}$  where  $y_\psi$  is some number in  $I_\psi$ , thus,  $y_\psi \sim \nu_\psi$ . Finally, given  $\psi$  has unit energy, i.e.  $A_\psi = N_\psi^{1/4}$

$$\text{var}\widehat{W}(s) \sim \frac{c_2}{N\nu_\psi^{\alpha+1}} s^\alpha \left(1 - \frac{s^\alpha c_2}{\nu_\psi^{\alpha+1} \sqrt{N_\psi}}\right) \quad (59)$$

For small scales ( $s \rightarrow 0$ ), the variance behaves like  $\text{var}\widehat{W}(s) \sim O(s^\alpha)$ . Figures 4.3(a)–(c) show empirical variance  $\text{var}\widehat{W}(s)$  varying with parameters  $N$ ,  $s$  and  $N_\psi$ , attesting the good agreement between experimental and theoretical results.

Let us now consider the new variable  $\log \widehat{W}(s)$ . With a central limit theorem argument, we can say that  $\widehat{W}(s)$  is asymptotically normal with mean  $\delta_s \approx As^\alpha$  and variance  $\sigma_s^2 \approx Cs^\alpha$ . Then, in first approximation, using a result on functions of asymptotically Gaussian variables [29, 22], we conclude that  $\log \widehat{W}(s)$  is asymptotically Gaussian and

$$\begin{cases} \mathbb{E} \log \widehat{W}(s) \approx \log \mathbb{E} \widehat{W}(s) \approx \log A + \alpha \cdot \log(s) \\ \text{var} \log \widehat{W}(s) \approx |\mathbb{E} \widehat{W}(s)|^{-2} \text{var} \widehat{W}(s) \approx B/A \cdot s^{-\alpha} \end{cases} \quad (60)$$

See figure 4.3(d).

To summarize the above, we propose to estimate the scaling exponent of the wavelet coefficients  $\rho^+$  via  $\widehat{\alpha}$ . For (asymptotically) Gaussian random variables such as (60), the maximum likelihood estimator of  $\alpha$  is simply obtained from a linear regression of  $\log \widehat{W}(s)$  against  $\log s$ , as already suggested in (44). Asymptotically, the resulting estimate  $\widehat{\alpha}$  converges to  $\rho^+$  defined by the lim sup of (27). In practice though, the finite size data set limits the regression range to some interval  $s \in (s_{\min} > 0, s_{\max} < \infty)$ . The important issue of properly choosing this scaling region is treated in the next section.

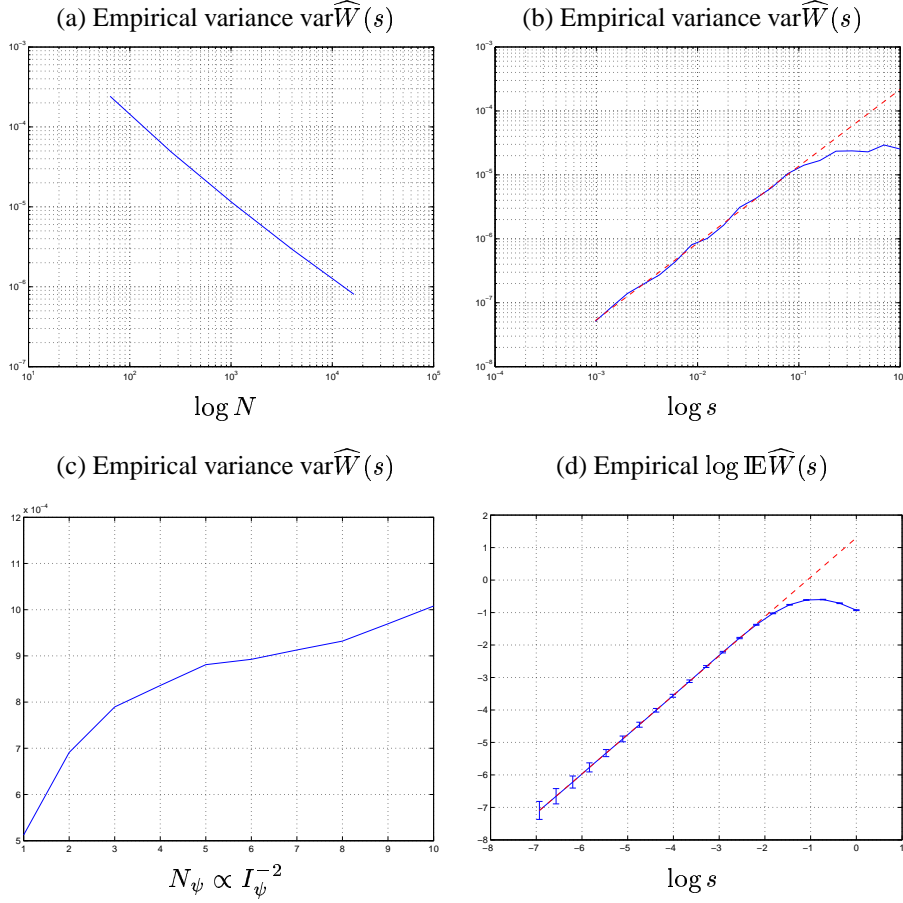


Figure 1: Experimental verification of expressions (59) and (60). (a)–(c): Empirical estimates of  $\text{var}\widehat{W}(s)$  estimated over a set of 100 independent realizations of  $\alpha$ -stable processes of length  $N$ . Evolution of  $\text{var}\widehat{W}(s)$  is plotted as a function of: (a)  $\log N$  ( $\alpha = 1.2$ ,  $N_\psi = 4$ ,  $s = 0.0087$ ); (b)  $\log s$  ( $\alpha = 1.2$ ,  $N_\psi = 4$ ,  $N = 2^{14}$ ); (c)  $N_\psi$  ( $\alpha = 1.2$ ,  $s = 0.0087$ ,  $N = 2^{14}$ ). (d) Empirical estimation of  $\log \mathbb{E}\widehat{W}(s)$  versus  $\log s$ . The error bars correspond to the standard deviation of  $\log \widehat{W}(s)$ . The dashed line materializes the theoretical law  $\log \mathbb{E}\widehat{W}(s) = \alpha \cdot \log(s) + C'$  ( $\alpha = 1.2$ ,  $N_\psi = 4$ ,  $N = 2^{14}$ ).

## 4.4 Choice of the scale range

We have defined an estimator for  $\rho^+$  via a log-log linear fit. While in theory the wavelet coefficients should decay as a powerlaw of the scale, we are in practice faced with the fact that the scaling deviates significantly from the theoretical ideal case for both large and small scales. Here we discuss the reasons for this deviation and explain how to choose the scaling region.

### 4.4.1 Lower bound of scaling region

We have two different approaches to determine a lower bound for the scale range of the linear regression  $\log \widehat{W}(s)$  versus  $\log s$  in (44).

The first one is based on a Shannon-like theorem. Our estimator estimates the singularity of the characteristic function at the origin. In practice, we use the empirical estimator for the characteristic function, i.e.,  $\hat{\phi}(u_k) = N^{-1} \sum_j \exp(iu_k X_j)$ . The maximum variation of  $\hat{\phi}$  is controlled by the maximum value of  $X_j$ . Therefore sampling  $\hat{\phi}$  at a higher rate than approximately  $(\overline{X})^{-1}$  with  $\overline{X} = \max\{X_j, j = 1, \dots, N\}$ , does not bring any finer information on the regularity of  $\phi(u)$  at  $u = 0$ . On the contrary, when the analyzing scale goes below the minimum bound  $s_{\min} = (\overline{X})^{-1}$ , the measured regularity is overestimated, as the function under analysis reduces to the sole  $C^\infty$  component  $\exp(iu_k \overline{X})$ , oversampled at the vicinity of the origin. Thus, concordantly with theorem 5 and lemma 7, when  $\hat{\alpha}$  is estimated from data below this minimum scale bound it reflects the regularity  $N_\psi$  of the analyzing wavelet rather than the targeted regularity of the characteristic function.

The second approach starts from the expression (43). In order to be consistent, we need to ensure that at least one sample  $X_j$  falls inside the equivalent support of  $\Psi(sx)$ . For small  $s$ , only the largest values of  $X_j$  are retained to enter the sum (43). As a result, if  $\overline{X}$  is the maximum sample of the series  $X_j$ ,  $\nu_\psi/s$  is the *central* frequency of the wavelet at scale  $s$ . We then want  $\overline{X} \approx \nu_\psi/s$ , which leads to  $s_{\min} \approx \nu_\psi/\overline{X}$ .

In summary, both arguments above lead to the same conclusion that the lower cut-off scale should be chosen proportionally to  $1/\overline{X}$ . For the numerical analysis in this paper we adopted:

$$s_{\min} = 1/\overline{X}. \quad (61)$$

Using a stable law with index of stability (or shape parameter)  $\alpha$ , we present in figure 4.4.2 the theoretical lower scale bound  $s_{\min} = (\overline{X})^{-1}$ . A linear regression of  $\log \widehat{W}(s)$  versus  $\log s$  for  $s > s_{\min}$  yields an accurate estimate of the characteristic exponent  $\rho^+ = \alpha$ . Moreover, on this same plot, we verify that for  $s < s_{\min}$ , the wavelet estimator  $\widehat{W}(s)$  behaves like  $s^{N_\psi}$ , in accordance with the aforementioned argument that the function under analysis is now the  $C^\infty$  exponential  $\exp(iu\overline{X})$ .

**Remark 19 (Lower cutoff for stable distribution)** For  $K$  independent samples of a  $\alpha$ -stable variable, the distribution of the maxima is such that

$$\begin{aligned} P[\max\{X_1 \cdots X_K\} \leq \nu_\psi/s] &= P[X \leq \nu_\psi/s]^K \\ &= (1 - C_\psi(s/\nu_\psi)^\alpha)^K \\ &\approx 1 - C_\psi K (s/\nu_\psi)^\alpha + O(s^{2\alpha}) \end{aligned}$$

We want  $P[\max\{X_1 \cdots X_K\} \geq \nu_\psi/s] \sim C_\psi K s^\alpha = 1$  (at least one sample!) which brings us for  $K$  independent stable samples (i.i.d. random variables) to

$$s_{\min} = C_\psi K^{-1/\alpha}. \quad (62)$$



#### 4.4.2 Upper bound and negative moments

As we saw, existence of moments is dictated by the tail decay of the distribution  $F(x)$  for  $x \rightarrow \infty$ . For instance, it is shown in [28], that the asymptotic tail behavior of a stable law is Pareto when  $0 < \alpha < 2$ . Defining when exactly this asymptotic behavior starts seems to be a tough problem (see [21]), as it depends heavily on the parameterization that is used to model the distribution (in the parametric context). We just pretend here, that the upper cutoff scale  $s_{\max}$  below which  $\widehat{W}(s)$  behaves like  $s^\alpha$  is also determined by this cutoff value of  $X$  separating the tail behavior (as a Pareto law for instance) from the *body* of the distribution. We illustrate this with a compound distribution, made out of a uniform distribution for  $|X| \leq \delta$  and  $\alpha$ -stable distribution for  $|X| > \delta$ . We show with this simple example (see Figure 4.4.2), that the upper cutoff scale is of order:

$$s_{\max} = \delta^{-1} \quad (63)$$

where  $\delta$  marks the transition from body to tail behavior in the distribution. In practical applications one might choose  $\delta$  from prior knowledge (rendering the estimator semi-parametric) or estimate  $\delta$  itself from the scaling plots (see Figure 4.4.2).

Beyond this upper limit, the wavelet estimator  $\log \widehat{W}(s)$  decays with slope  $-1$ . This particular value of the slope depends only upon the distribution we have chosen for the body of our compound distribution. In our example, the uniform distribution has negative moments only for  $p > \lambda^-(X) = -1$ . That is precisely this bound that is estimated in 4.4.2, when  $s > s_{\max}$ . The same value would have been estimated, if instead of  $X$  directly we had analyzed the new random variable  $Y = X^{-1}$  as discussed in Section 3.4, and estimated  $\lambda^+(Y) = -\lambda^-(X)$  from the tail decay of the transformed distribution.

This observation bears a convenient consequence as far as negative moments are concerned: We can fully exploit the behavior of  $\widehat{W}(s)$  for  $s > s_{\max}$ , leading us to a simple estimator of  $\lambda^-$  in (1). To illustrate this, we now choose a compound process similar to before but replace the uniform distribution for  $|X| \leq \delta$  with a Gamma distribution of parameter  $0 < \gamma < 1$ . The density of the Gamma distribution behaves as  $|\cdot|^{\gamma+1}$  around the origin, whence negative moments exist exactly for negative orders  $p > \lambda^- = -\gamma$ . Therefore, we estimate the slope of  $\log \widehat{W}(s)$  for  $s > s_{\max} = \delta^{-1}$  and compare this estimate against the theoretical value  $\lambda^- = -\gamma$  (see table 2).

To summarize, given  $K$  i.i.d. random variables  $\{X_j, j = 1, \dots, K\}$ , the wavelet estimator (43) behaves like:

- $\widehat{W}(s) \sim s^{N_\psi}$ , for  $s < s_{\min} = (\max_j \{X_j\})^{-1}$ ,
- $\widehat{W}(s) \sim s^{\rho^+}$ , for  $s_{\min} < s < s_{\max}$ , where  $s_{\max}$  corresponds to the inverse of the cut-off value separating the tail from the body of the underlying distribution,

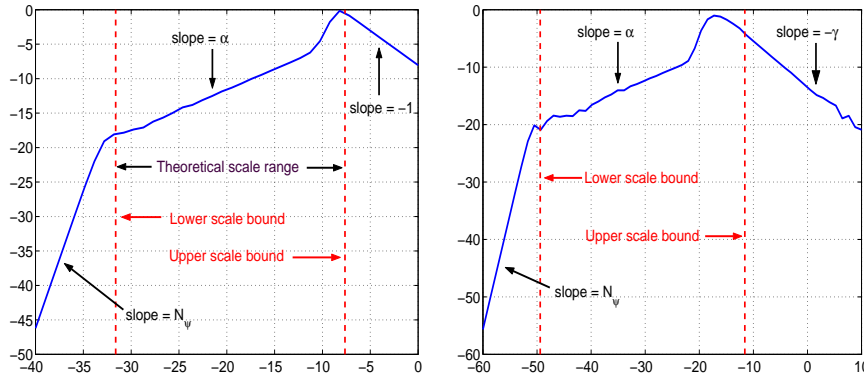


Figure 2: SCALING REGION AND CUTOFF SCALES: Choosing the scale too small, the resolution is fine enough for the wavelet to analyze the individual exponentials that form the estimator  $\hat{\phi}$ . According to lemma 7 the wavelet coefficients decay (at least) with exponent  $N_\psi$ . Choosing the scale too large, the estimator samples the body instead of the tail of the distribution; thus, the wavelet coefficients adhere to a powerlaw with exponent  $\lambda^-$ .

- $\widehat{W}(s) \sim s^{\rho^-}$ , for  $s > s_{\max}$ .

This is impressively demonstrated in Figure 4.4.2. In fine, both  $\rho^+$  and  $\rho^-$  can be deduced from a linear regression of  $\log \widehat{W}(s)$  versus  $\log s$ , over the corresponding scale intervals. As elaborated in section 3 choosing a wavelet according to theorem 16 and corollary 17 we have  $\lambda^+ = \rho^+$  and  $\lambda^- = \rho^-$  whenever these numbers are smaller than 2; in general,  $\rho^+ \leq \lambda^+ \leq \lfloor \rho^+ \rfloor + 1$  and similar for  $\lambda^-$ .

#### 4.5 Choice of the wavelet

The theoretical results of section 3 form the basis of our estimator. For them to hold the analyzing wavelet  $\psi$  is required to have a semi-definite Fourier transform as well as a number of vanishing moments  $N_\psi$  larger than  $\mathcal{H}_{\text{Re } \phi}(0)$ .

In practice, we suggest to start with a low regularity wavelet such as the second derivative of the gaussian window  $\psi_2(t)$ , corresponding to  $N_\psi = 2$ . If the slope  $\rho^+$  obtained from the linear regression of  $\log \widehat{W}(s)$  versus  $\log s$  is smaller than  $N_\psi = 2$ , then theorem 11 immediately posits that the positive critical order  $\lambda^+$  is equal to  $\rho^+$ . Now, if the measured slope  $\rho^+$  equals  $N_\psi = 2$ , we need to verify whether the regularity  $\mathcal{H}_{\text{Re } \phi}(0)$  is actually larger than two or not.

To this end, we increase the number of vanishing moments  $N_\psi = p$  of  $\psi_p(t)$ , and repeat the estimation of  $\rho^+$  for increasing integer  $p$  as long as the slope  $\rho^+$  hits the bound  $N_\psi$ . Once we get a  $\rho^+ < N_\psi$ , we should recall corollary 17 which only guarantees that  $\lambda^+$  can not exceed  $\lfloor \rho^+ \rfloor + 1$ . This could be of interest in itself for model verification.

To obtain optimal results, in a second step, we must turn to lemma 14 to check a posteriori that indeed  $\lambda^+ = \rho^+$ . Rigorously, the procedure should be as follows : considering Gaussian derivatives

of fractional order  $\eta \in (\lfloor \rho^+ \rfloor, \lfloor \rho^+ \rfloor + 1)$ , we form the ratio  $\widehat{W}_\eta(s)/s^\eta$ , and study its limit when  $s$  goes towards zero. When it is finite, this limit corresponds to  $\mathcal{M}_\eta$ , the absolute moment of order  $\eta$  of r.v.  $X$ . Finally, the critical exponent  $\lambda^+$  is then determined by the frontier :

$$\begin{aligned} \lambda^+ &= \sup_{\eta} \left\{ \eta \in (\lfloor \rho^+ \rfloor, \lfloor \rho^+ \rfloor + 1) : \limsup_{s \rightarrow 0} \widehat{W}_\eta(s)/s^\eta < +\infty \right\} \\ &= \inf_{\eta} \left\{ \eta \in (\lfloor \rho^+ \rfloor, \lfloor \rho^+ \rfloor + 1) : \limsup_{s \rightarrow 0} \widehat{W}_\eta(s)/s^\eta = +\infty \right\}. \end{aligned} \quad (64)$$

In practice though, determining this frontier is not quite realistic, and the reason is twofold. First, as we saw, finite size data sets do not allow for analyzing scales close to zero, and limiting the ratio in (64) to  $s \leq s_{\min}$ , loosens considerably the breakpoint separating convergence from divergence. Second, recalling result of proposition 18, once  $\eta$  is too close to the distribution tail decay  $\alpha$ , the approximation  $\mathbb{E}\widehat{W}(s) \sim A \cdot s^\alpha$ ,  $s \rightarrow 0$ , does not hold anymore. The residual term  $B \cdot s^\eta$  may not be negligible, and can introduce some bias on the estimate of  $\alpha$ .

Despite all this, in all our experiments, we observed that the *basic* estimate  $\rho^+$  obtained with  $N_\psi > \mathcal{H}_{\text{Re } \phi}(0)$  already accurately estimates  $\lambda^+$  on its own. In particular, we never encountered the case  $\rho^+ < \lambda^+ < \lfloor \rho^+ \rfloor + 1$ , that indeed, would necessitate the above, more refined procedure to identify the characteristic exponent precisely.

## 5 Applications

Application of particular interest in this context are the parameter estimation of stable laws as well as the estimation of the multifractal partition function.

### 5.1 Estimating Stable and Gamma Parameters

To set notation we recall some classes of distributions well known in the literature, that we will use to illustrate our characteristic exponent estimator.

**Pareto.** A Pareto density  $p_X$  is a simple power law function that take on the form

$$p_X(x) = \begin{cases} \alpha \mu^\alpha x^{-\alpha-1} & \text{if } x > \mu, \\ 0 & \text{else,} \end{cases} \quad (65)$$

with  $\alpha$  the shape parameter, and  $\mu$  the position parameter. A random variable  $X$  with Pareto distribution, has positive  $r$ -th order moments existing only for orders  $r < \alpha$ , while all negative orders moments exist as  $X \geq \alpha > 0$  almost surely. The median is  $\mu 2^{1/\alpha}$ , and if  $\alpha > 1$  then the mean exists and equals  $\mathbb{E}X = \mu\alpha/(\alpha - 1)$ .

**Stable.** Stable laws form a class of heavy tailed distributions, for which there exists an abundant literature (see e.g. [28] for a detailed introduction). A random variable  $X$  follows a stable law that we denote  $S\alpha S(\sigma, \beta, \mu)$ , if and only if its characteristic function reads :

$$\mathbb{E}[\exp(iuX)] = \exp(-\sigma^\alpha |u|^\alpha (1 - i\beta w_\alpha(t)) + i\mu u), \quad (66)$$

where  $w_\alpha(t) = \tan(\pi\alpha \operatorname{sgn}(t)/2)$  for  $\alpha \neq 1$  and  $w_1(t) = -(2/\pi) \operatorname{sgn}(t) \log |t|$ .

Although there exists no closed form for stable distributions except for a handful of special cases, stable laws have a tail behavior that can be approximated as a Pareto distribution (65). Indeed, [28, Property 1.2.15] reads as: If  $X \sim S\alpha S(\sigma, \beta, \mu)$  with  $0 < \alpha < 2$ , then

$$\begin{aligned} \lim_{\lambda \rightarrow \infty} \lambda^\alpha P[X > \lambda] &= C_\alpha \frac{1+\beta}{2} \sigma^\alpha \\ \lim_{\lambda \rightarrow \infty} \lambda^\alpha P[X < -\lambda] &= C_\alpha \frac{1-\beta}{2} \sigma^\alpha \end{aligned} \quad (67)$$

where  $1/C_\alpha = \int_0^\infty x^{-\alpha} \sin(x) dx$  depends only on  $\alpha$ .

The index  $\alpha$  is sometimes referred to as the characteristic exponent of the stable law, and for our purpose, it constitutes the most important parameter since absolute moments of order  $r$  are finite exactly for  $r \in (-1, \alpha)$  ( $0 < \alpha < 2$ ). For  $\alpha = 2$  we recover the special case of Gaussian distribution, with existing moments at all orders  $r > -1$ . The parameter  $\sigma$  indicates scale, since  $X \sim S\alpha S(\sigma, \beta, \mu)$ , then  $aX \sim S\alpha S(a\sigma, \beta, a\mu)$  ( $a > 0$ ). For  $\alpha = 2$  we have  $\sigma^2 = \operatorname{var}(X)/2$  while for  $\alpha < 2$  the second moment  $\mathbb{E}[X^2]$  is infinite and the variance is not defined. The parameter  $\mu$  defines position in the sense that if  $X \sim S\alpha S(\sigma, \beta, \mu)$  then  $X + c \sim S\alpha S(\sigma, \beta, \mu + c)$ . Provided that  $\alpha > 1$  we may be even more explicit and identify  $\mu$  as the expected value:  $\mathbb{E}[X] = \mu$ . However, in the case  $\alpha \leq 1$  the mean  $\mathbb{E}[X]$  is not even defined; as the most prominent example we mention the Cauchy distribution. Finally, the parameter  $\beta$  provides a measure for the skew, more precisely,  $X$  is symmetrical if and only if  $\beta = 0$ ; moreover, if this is the case then (66) reduces to (2).

**Gamma.** The last case we will comment on is the Gamma distribution. A random variable  $X$  has Gamma distribution if

$$p_X(x) = \begin{cases} \lambda x^{\gamma-1} \exp\{-cx\} & \text{if } x \geq 0 \\ 0 & \text{else.} \end{cases} \quad (68)$$

In the above,  $\gamma$  and  $c$  are positive numbers, and  $\lambda = c^\gamma / \Gamma(\gamma)$ , with  $\Gamma$  the generalized factorial function. The special case  $\gamma = n/2$ ,  $c = 1/2$  with  $n$  an integer, corresponds to the Chi-square density with  $n$  degrees of freedom, and for  $n = 2$  it reduces to the usual exponential density. As far as moments are concerned, thanks to the dominant exponential decay in (68), all positive order moments exist, and in particular  $\mathbb{E}X = \gamma/c$  and  $\mathbb{E}X^2 = \gamma(\gamma + 1)/c^2$ . The negative moments, i.e.,

$$\mathcal{M}_r = \int_0^\infty \lambda x^{r+\gamma-1} \exp\{-cx\} dx, \quad r < 0, \quad (69)$$

converge only for  $r > -\gamma$ .



Table 1: Estimation of the characteristic exponent  $\alpha$  of a stable law, using Koutrouvelis procedure, McCulloch procedure and our wavelet based procedure, using  $N = 2^{12}$  i.i.d. samples of a stable variable. Scale, position and skew parameters are fixed ( $\sigma = 1, \mu = 0, \beta = 0$ ), and  $\alpha$  varies in  $(0, 2)$ . Empirical standard deviations on the estimates are based upon a 1000 realizations set.

| $\alpha$                          | 0.2               | 0.6               | 1               | 1.4               | 1.8              |
|-----------------------------------|-------------------|-------------------|-----------------|-------------------|------------------|
| $\widehat{\lambda}^+$             | $0.196 \pm 0.007$ | $0.58 \pm 0.018$  | $1.0 \pm 0.035$ | $1.46 \pm 0.066$  | $1.74 \pm 0.02$  |
| $\widehat{\alpha}$ (Koutrouvelis) | ND                | $0.60 \pm 0.007$  | $1.0 \pm 0.009$ | $1.403 \pm 0.013$ | $1.80 \pm 0.012$ |
| $\widehat{\alpha}$ (McCulloch)    | $0.59 \pm 0.0018$ | $0.605 \pm 0.009$ | $1.0 \pm 0.009$ | $1.40 \pm 0.016$  | $1.80 \pm 0.022$ |

For the above classes of distributions, Pareto, stable and Gamma, there exist efficient procedures aimed at estimating the different sets of parameters. In most cases, these estimators are parametric estimators and they turn out to be optimal (in the sense of maximum likelihood) whenever the specific underlying distribution model and the analyzed data distribution do match. Our estimator (43) is non-parametric, and it should not be expected to outperform a parametric estimator on the distribution it is tailored for. This is for instance very clear on the experiments depicted in Table 1. Considering  $N$  i.i.d. samples of a stable variable  $X \sim S_{\alpha}S(\sigma, \beta, \mu)$ , we compare our estimates (44) of  $\alpha$  against two well-known parametric estimators for stable laws: Koutrouvelis [16] and McCulloch [19] procedures.

Superiority of parametric estimators in this appropriate context is not questionable. However, in most real world applications, the true density underlying the data to be analyzed is rarely known, and very likely blind application of parametric estimators will produce aberrant results. A very illustrative example is proposed in Table 2. We consider a Gamma variable  $X$  with shape parameter  $0 < \gamma < 1$ , and form the new variable  $Y = X^{-1}$ . From (69) we know that  $r$ -th order moments of  $Y$  should only exist for  $r < \gamma$ . If we now compare the (empirical) densities derived both from  $Y$  and from a stable variable with characteristic exponent  $\alpha = \gamma$  and skewness parameter  $\beta = 1$  (which ensures positivity since  $\alpha < 1$ ) it is quite difficult to dissociate them (Figure 3). Yet, applying crudely stable law designed estimators, like Koutrouvelis or McCulloch, to the raw data  $Y$ , yields very bad estimates  $\widehat{\alpha} = \widehat{\gamma} = -\widehat{\lambda}^-$ . In contrast, determining the characteristic exponent  $\lambda^+(Y) = -\lambda^-(X)$  from our wavelet-based regression procedure (as described in Section 3.4), provides us with fairly good estimates of shape parameters  $\gamma$  for Gamma distributions. Hence, because our non-parametric estimator does not assume any a priori distribution for the data, it compares favorably as a general purpose tool to parametric estimators.

## 5.2 Estimating the Multifractal partition function

Processes with continuous but highly irregular paths occur quite commonly in a host of applications [2, 17, 26]. For cascading measures, e.g, such discontinuities convey most of the pertinent information and their relevance is widely agreed upon. More generally, various complex systems are known to produce signals in which singularities may coexist in great multitude and in a highly interwoven

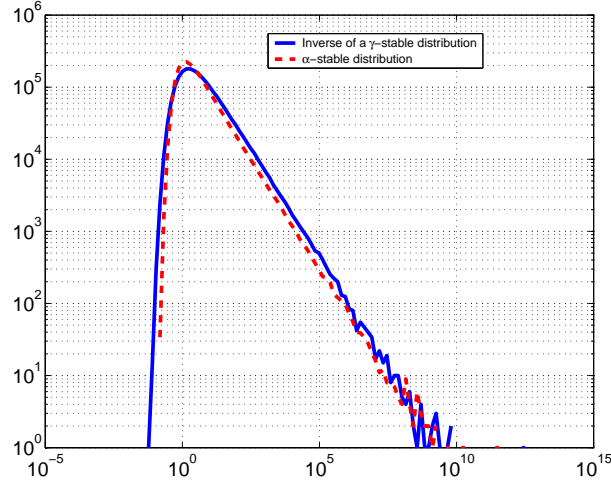


Figure 3: Empirical distributions of the random variables  $Y = X^{-1}$  and  $Z$ , where  $X$  follows a Gamma law with  $\gamma = 0.6$ , and  $Z$  follows a stable law with  $\alpha = 0.6$  and  $\beta = 1$ . For both cases,  $\lambda^+ = \alpha = \gamma$ . Axis are in logarithmic scale.

Table 2: Estimation of the shape exponent  $\gamma$  from a Gamma variable  $X$ . Koutrouvelis procedure, McCulloch procedure and our wavelet based procedure are applied to the heavy tail transformed variable  $Y = X^{-1}$ .  $N = 2^{12}$  i.i.d. samples of Gamma variable where used. Parameters  $c = 1$  is fixed, and  $\gamma$  varies in  $(0, 1)$ . Empirical standard deviations on the estimates are based upon a 1000 realizations set.

| $\gamma$                      | 0.2               | 0.4               | 0.6               | 0.8              |
|-------------------------------|-------------------|-------------------|-------------------|------------------|
| $-\widehat{\lambda^-}$        | $0.204 \pm 0.007$ | $0.395 \pm 0.008$ | $0.589 \pm 0.015$ | $0.793 \pm 0.03$ |
| $\hat{\alpha}$ (Koutrouvelis) | ND                | $0.433 \pm 0.006$ | $0.56 \pm 0.007$  | $0.67 \pm 0.009$ |
| $\hat{\alpha}$ (McCulloch)    | $0.513 \pm 0.000$ | $0.514 \pm 0.000$ | $0.583 \pm 0.009$ | $0.72 \pm 0.013$ |

interplay. The local degree of Hölder regularity  $\mathcal{H}(t)$  becomes then so utterly erratic as a function of location  $t$  that its pointwise estimation via definition 4 becomes completely unrealistic.

With the estimation of  $\mathcal{H}(t)$  point by point infeasible one resorts to the "multifractal formalism" which allows to quantify how frequently a given singularity strength  $\mathcal{H}(t) = h$  is assumed, where frequency can be measured in geometrical or probabilistic terms (as the location  $t$  is chosen randomly). Consequently, several notions of so-called multifractal spectra have been proposed (see for instance the introduction of [24] for a list of a variety of approaches) to quantify this occurrence in various terms. For simplicity, we will focus here on a wavelet-based multifractal partition function.

For a random process  $X(t)$ , the *partition function*  $\tau(q)$  is defined as<sup>4</sup>

$$\tau(q) \triangleq \lim_{s \rightarrow 0^+} \frac{\log \mathbb{E} \left[ |W_X(t, s)|^q \right]}{\log s}, \quad (70)$$

where  $W[X](t, s)$  is a wavelet decomposition (13) of  $X$ . The partition function measures (i) the power law scaling of the moments (ii) higher-order dependencies of the wavelet coefficients and (iii) the singularity structure of the process, all in one. Note that  $\tau(q)$  is always *concave*, since moment generating functions are log-convex.

As a means to single out processes with highly erratic Hölder regularity  $\mathcal{H}(t)$  we refer to a process  $X$  as a *multifractal* if its multifractal partition function is non-linear. It is known from multifractal theory that an honestly concave (non-linear) partition function is indicative of the presence of a rich range of different degrees of Hölder regularity in the signal which occur in a highly erratic way (see, e.g., [27]). To the contrary, a (mono-)fractal process which spots a smoothly varying degree of regularity, possesses a piecewise linear partition function. In particular, if  $\mathcal{H}(t) = H, \forall t$  then  $\tau(q) = qH$ .

A great difficulty and source of inaccuracy in multifractal analysis stems from infinite moments: to an innocent empirical estimator of moments, divergence might not be apparent. Consequently, a numerical procedure might return finite estimates of  $\tau(q)$  from (70) even if the moment of order  $q$  does not exist. As we are about to demonstrate with the simple examples of self-similar processes, this estimation inaccuracy may lead to a concave *estimate* of  $\tau(q)$ , although the actual process is mono-fractal with a *linear* partition function.

In order to distinguish multifractal from mono-fractal processes, we must, therefore, restrict the support of  $\tau(q)$  to the actual range of orders  $q$ — with finite moments  $\mathbb{E}|W[X](t, s)|^q$ .

We now present some examples of well-studied processes commonly used as fractal and multifractal models.

### Fractional Brownian motion

**Definition 20** A fractional Brownian motion  $B_H(t)$  with index  $0 < H < 1$  (noted  $H$ -fBm thereafter) is defined as the unique process that verifies simultaneously the following conditions [7]:

1. with probability one  $B_H(t)$  is continuous and  $B_H(0) = 0$

---

<sup>4</sup>An earlier version of  $\tau(q)$  relies on the  $q$ -th order moments of the increment process  $Y_s(t) \triangleq X(t + st_0) - X(t)$  instead of the wavelet coefficients  $W[X](t, s)$ .

2. for any  $t \geq 0$  and  $\delta > 0$ , the increment process  $y^{(\delta)}(t) \triangleq B_H(t + \delta) - B_H(t)$  is stationary and has normal distribution with zero mean and variance  $\delta^{2H}$ .

The index  $H$  controls the regularity of the graph  $(t, B_H(t))$  with  $\mathcal{H}(t) = H$  for almost all  $t$ . As a result,  $H$ -fBm is a mono-fractal Gaussian process. Moreover, it is known that its wavelet decomposition  $W[B_H](t, s)$  forms a stationary Gaussian process at each scale  $s$  [11]. Finally, the multifractal partition function (70) is linear and reads :

$$\tau(q) = \begin{cases} qH & \text{for } q > -1, \\ -\infty & \text{else.} \end{cases} \quad (71)$$

**Self-Similar stable processes.** Self-similar stable processes are obtained when replacing normal by stable distributions in condition (2) of the above definition of  $H$ -fBm. Such processes have discontinuous paths with interesting fractal properties and they are well adapted for infinite variance stochastic modelling. A rigorous definition reads as follows: [28]

$$X(t) \triangleq \int f(t, u) M(du). \quad (72)$$

In this integral representation,  $M$  is a symmetric stable measure with characteristic exponent  $\alpha$  and  $f(t, u)$  is an integral kernel that controls the fractal properties of  $X$ . In particular, choosing  $f(t, u) = (t - u)_+^{H-1/\gamma} - (-u)_+^{H-1/\gamma}$ , with  $(u)_+ = \max(u, 0)$ ,  $\gamma \in (0, 2)$  and  $H \neq 1/\gamma \in (0, 1)$ , yields the *linear fractional stable motion* [28]. A detailed study of this process [6, 23] showed that at each scale, the wavelet coefficients of its decomposition are identically distributed with a stable law of same index  $\alpha$  as  $X$ . Then, taking into account the range of existence ( $\lambda^- = -1 < q < \alpha = \lambda^+$ ) for the moments of a stable random variable and following the lines of  $H$ -fBm analysis, we get [27]:

$$\tau(q) = \begin{cases} qH & \text{for } -1 < q < \alpha, \\ -\infty & \text{else.} \end{cases} \quad (73)$$

Here again,  $\tau(q)$  is linear over the range  $q \in (\lambda^-, \lambda^+)$  of finite moments  $\mathbb{E}|W[X](t, s)|^q$ .

**Multifractal processes.** *Mono-fractals* like  $H$ -fBm and self-similar stable processes, albeit providing us with elegant fractal paradigms, often lack flexibility to serve as models in real-world situations, due to their simplicity in scaling or due to non-stationarities. To meet the need for closed form models with simple non-linear partition functions yet stationary, zero mean increments the following multifractal processes were proposed in [27], based on an idea of B. Mandelbrot:

$$\mathcal{B}(t) \triangleq B_H(\mathcal{M}(t)). \quad (74)$$

In this definition  $\mathcal{M}(t)$  stands for a monotonic multiplicative cascade with flexible multifractal properties (i.e. non-linear multifractal partition function  $\tau_{\mathcal{M}}(q)$ ), and  $B_H$  represents an  $H$ -fBm (or more generally any self-similar process with index  $H$ ). Based on a multifractal time warping,  $\mathcal{B}(t)$  combines in one the rich multifractal structure of a multiplicative cascade along with the self-similarity

and the non-monotonicity of a  $H$ -fBm paths. For this compound process we have the fundamental result [27]:

$$\tau_{\mathcal{B}}(q) = \begin{cases} \tau_{\mathcal{M}}(qH) & \text{for } q > -1 \\ -\infty & \text{else.} \end{cases} \quad (75)$$

This partition function is clearly non-linear as it inherits from the multifractal structure of the time measure  $\mathcal{M}(t)$ .

**Empirical evidence for multifractals** Using the self-similar stable process (72) and the time warped fBm (74) as paradigms for fractal and multifractal processes respectively, we now show how ambiguous it can be to decide between true multi-scaling and mono-scaling behaviors from linearity of the empirical partition function.

In [4], a thorough study on self-similar stable processes stresses that empirical estimators of  $\mathbb{E}|W[X](t, s)|^q$ , based on a finite length data set, are not sensitive to theoretical divergence of moments. More precisely, whereas  $q$ -th order moments of a self-similar stable process should not exist for  $q \notin (-1, \alpha)$ , there is no clear evidence of this divergence in practice. Moreover, the non-linearity of the logarithmic transform used in (70), creates an artificial scaling  $\mathbb{E} \log \hat{\mathbb{E}}|W[X](t, s)|^q \sim \hat{\tau}(q) \log s$  for  $q > \alpha$ , very akin to the ones occurring for  $q \in (-1, \alpha)$ . As a result, it seems completely legitimate in practice, to stretch out  $\tau(q)$  outside its theoretical support and get the non-linear *concave empirical partition function* for self-similar stable processes:

$$\hat{\tau}(q) = \begin{cases} qH & \text{for } q \in (-1, \alpha) \\ 1 + q(H - \frac{1}{\alpha}) & \text{for } q > \alpha. \end{cases} \quad (76)$$

In Figure 4, the power law scalings observed for the empirical estimates  $\hat{\mathbb{E}}|W[X](t, s)|^q$ , are plotted for both  $q \in (-1, \alpha)$  and  $q \notin (-1, \alpha)$ . In the same figure, the corresponding empirical multifractal partition function is clearly non-linear over the real line, and prompts to conclude that process  $X$  is multifractal.

The same analysis applied to the multifractal process  $\mathcal{B}(t)$  of definition (74) yields an empirical partition function that clearly deviates from linearity as well (Figure 5).

- In summary, these two trivial examples emphasize how misleading it is to simply rely on linearity of the partition function, to decide whether a process is fractal or multifractal.

To be fully reliable this linearity criterion must focus on the true support of  $\tau(q)$ , which is uniquely determined by the range  $q \in (\lambda^-, \lambda^+)$  of converging moments: that is where our wavelet-based characteristic exponent estimator (44) comes in as an intermediate ingredient of fractal analysis.

Back to the self-similar stable process  $X$  of (72), we know from [4], that for any fixed scale  $s_0$ , its wavelet decomposition  $W[X](t, s_0)$  has stable distribution with index  $\alpha$  imposed by the integrant symmetric stable measure  $M$  used in (72). We then simply run the procedure described in section 3 and section 4 on  $W[X](t, s_0)$ , and get the characteristic interval  $(\lambda^-, \lambda^+)$  on which  $\tau(q)$  is theoretically defined. Doing so, we find for our self-similar stable example, that the empirical

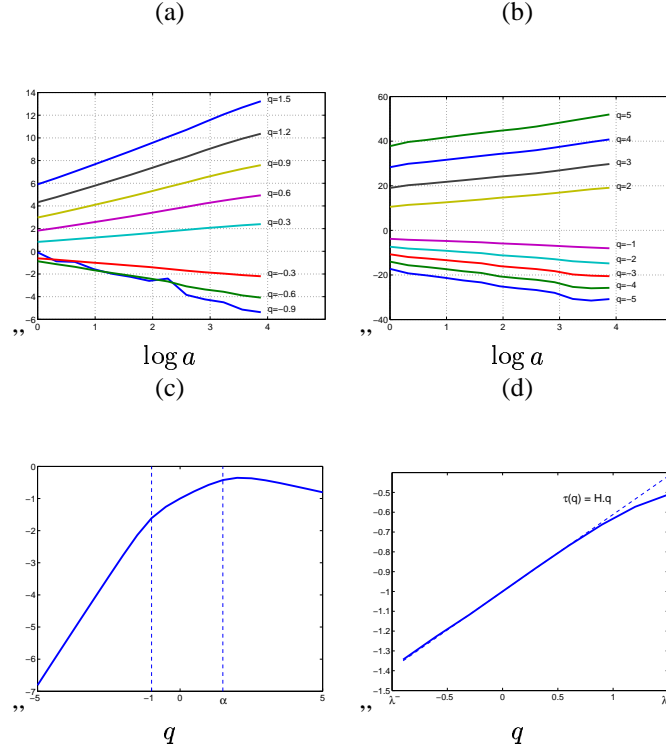


Figure 4: Multifractal analysis of a self-similar stable process (72) with  $\alpha = 1.5$ ,  $H = 0.47$ . We use a  $N = 2^{13}$  sample points realization. (a)  $\log \mathbb{E}|W[X](t, s)|^q$  versus  $\log s$  for  $q = -0.9$  (bottom curve) to  $q = \alpha$  (top curve). Within this range of  $q$ 's, log-moments scale linearly with slope  $\tau(q) = qH$ . (b)  $\log \mathbb{E}|W[X](t, s)|^q$  versus  $\log s$  for  $q \notin (-1, \alpha)$ . While in theory they should diverge, due to finite sample size effect and to non-linearity of the log function, log-moments keep scaling linearly with a different slope  $\tau(q) = 1 + q(H - 1/\alpha)$ ,  $q > \alpha$ . (c) Empirical partition function  $\hat{\tau}(q)$  computed over a non supervised interval  $(-5, 5)$ . (d) Empirical partition function  $\hat{\tau}(q)$  restricted to the predetermined characteristic interval  $(\hat{\lambda}^- = -1, \hat{\lambda}^+ = \alpha)$ .

estimate  $\hat{\tau}(q)$  restricted now to the predetermined interval  $(\hat{\lambda}^- = -1, \hat{\lambda}^+ = \alpha)$  displays a linear behavior coherent with the mono-fractality of the underlying process (Figure 4).

Regarding our multifractal example  $\mathcal{B}(t)$ , the normality of the process transposes to its wavelet decomposition  $W[\mathcal{B}](t, s_0)$ , meaning that all moments  $\mathbb{E}|W[\mathcal{B}](t, s_0)|^q$  exist for  $q \in (-1, +\infty)$ . In this particular case, restricting the empirical partition function to the characteristic interval  $(\hat{\lambda}^- = -1, \hat{\lambda}^+ = +\infty)$  does not linearize  $\hat{\tau}(q)$ , which remains the footprint of  $\mathcal{B}(t)$  multifractality (Figure 5).

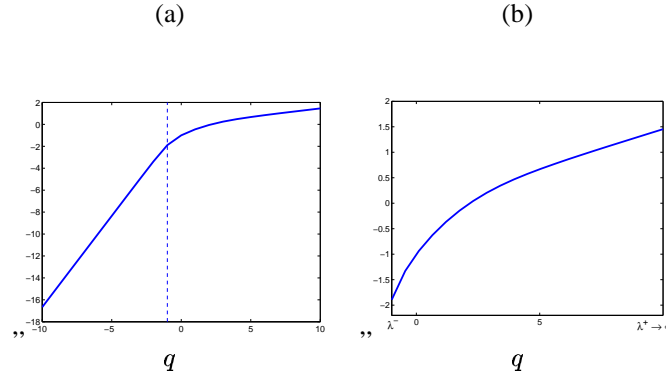


Figure 5: Multifractal analysis of a  $H$ -fBm in multifractal time (74) with  $H = 0.47$ , and  $\mathcal{M}$  a binomial multiplicative cascade. (a) Empirical partition function  $\hat{\tau}(q)$  computed over a non supervised interval  $(-5, 5)$ . (b) Empirical partition function  $\hat{\tau}(q)$  restricted to the predetermined characteristic interval  $(\hat{\lambda}^- = -1, \hat{\lambda}^+ \rightarrow \infty)$ .

## Discussion and Conclusions

We itemize the three main results we have derived in this paper.

- We have established a theoretical connection between what we called the characteristic exponent fixing the highest order  $\lambda^+$  of existing moments for a random variable, the tail decay of its probability distribution and the Holder regularity of its characteristic function at origin. The proposed theorem unifies and extends to any high orders, some known results that hold for  $\lambda^+ \leq 2$ .
- We proposed a wavelet based estimator of  $\lambda^+$  and  $\lambda^-$ , that in fine reduces to an extraordinarily simple implementation. Moreover, this characteristic exponent estimator is non-parametric and therefore does not assume any a priori knowledge on the underlying distribution.
- From an application point of view, this estimator shows very useful at characterizing rare events (often responsible for divergence of moments) and measuring power law decays of fat tail distributions. We also stressed a particularly interesting application of this estimator in the context of multifractal analysis, proving on a typical example how we could take advantage over the determination of  $\lambda^+$  and  $\lambda^-$ , to distinguish between fractal and multifractal processes.

## References

- [1] P. Abry, P. Gonçalvès, and P. Flandrin. Wavelets, spectrum analysis and  $1/f$  processes. In A. Antoniadis and G. Oppenheim, editors, *Lecture Notes in Statistics: Wavelets and Statistics*,

- volume 103, pages 15–29, 1995.
- [2] E Bacry, J.-F. Muzy, and A. Arneodo. Singularity spectrum of fractal signals from wavelet analysis: Exact results. *J. Stat. Phys.*, 70:635–674, 1993.
  - [3] K. G. Binmore and H. H. Stratton. A note on characteristic functions. *Ann. Math. Statist.*, 40:303–307, 1969.
  - [4] P. Chainais, P. Abry, and D. Veitch. Multifractal analysis and  $\alpha$ -stable processes: a methodological contribution. In *ICASSP*, Istanbul (Turkey), 2000.
  - [5] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, New York, 1992.
  - [6] L. Delbeke and P. Abry. Stochastic integral representation and properties of the wavelet coefficients of linear fractional stable motion. Preprint.
  - [7] Kenneth J. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley and Sons, New York, 1990.
  - [8] A Feuerverger and P. McDunnough. On some fourier methods for inference. *Journal of the American Statistical Association*, 76(374):379–387, 1981.
  - [9] A Feuerverger and P. McDunnough. On the efficiency of empirical characteristic function procedures. *J. R. Statist. Soc. B*, 43(1):20–27, 1981.
  - [10] A Feuerverger and R. A. Mureika. The empirical characteristic function and its applications. *Ann. Statist.*, 55:88–97, 1977.
  - [11] P. Flandrin. Wavelet analysis and synthesis of fractional brownian motion. *IEEE Trans. Info. Theory*, 38:910–917, 1992.
  - [12] S. Jaffard. Exposants de hölder en des points donnés et coefficients d’ondelettes. *C. R. Acad. Sc.*, 308:79–81, 1989.
  - [13] S. Jaffard. Local behavior of Riemann’s function. *Contemporary Mathematics*, 189:287–307, 1995.
  - [14] S. Jaffard. Méthodes d’ondelettes pour l’analyse multifractale de fonctions. In P. Abry, P. Gonçalves, and J. Lévy-Véhel, editors, *Fractals et lois d’échelle*, Traié Information–Commande–Communication. Hermès Sc. Pub., Paris, France, 2001.
  - [15] T. Kawata. *Fourier Analysis in Probability Theory*. Probability and Mathematical Statistics. Academic Press, 1972.
  - [16] I. A. Koutrouvelis. Regression-type estimation of the parameter of stable law. *J. American Statistical Association*, 75:918–928, 1980.
  - [17] W. Leland, M. Taqqu, W. Willinger, , and D. Wilson. On the self-similar nature of ethernet traffic. *IEEE, ACM Trans. Networking*, pages 1–15, 1994.



- [18] S. Mallat and W. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. on Information Theory*, 38(2), 1992.
- [19] J. H. McCulloch. Simple consistent estimator of stable distribution parameter. *Communication on Statistical Simulation*, 15:1109–1136, 1986.
- [20] J. H. McCulloch. Measuring tail thickness to estimate the stable index alpha: A critique. *American Statistical Association*, 15:74–81, 1997.
- [21] J. P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. preprint.
- [22] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. Mc Graw Hill, 3 edition, 1991.
- [23] B. Pesquet-Popescu. Statistical properties of the wavelet decomposition of certain non-gaussian self-similar processes. *IEEE Trans. on Sig. Proc.*, 75:303–322, 1999.
- [24] R. Riedi. An improved multifractal formalism and self-similar measures. *J. Math Ana. Appl.*, 189:462–490, 1995.
- [25] R. Riedi. Lois d'échelle multifractales: fondements et approche ondelettes. In P. Abry, P. Gonçalvès, and J. Lévy-Véhel, editors, *Fractals et lois d'échelle*, Traité Information–Commande–Communication. Hermès Sc. Pub., Paris, France, 2001.
- [26] R. Riedi, M. S. Crouse, V. Ribeiro, and R. G. Baraniuk. A multifractal wavelet model with application to TCP network traffic. *IEEE Trans. Info. Theory*, 1999. Special issue on multiscale statistical signal analysis and its applications.
- [27] R. H. Riedi. Multifractal processes. in: “Long range dependence : theory and applications”, eds. Doukhan, Oppenheim and Taqqu, to appear 2002.
- [28] G. Samorodnitsky and M. Taqqu. *Stable non-Gaussian random processes*. Chapman and Hall, New York ISBN 0-412-05171-0, 1994.
- [29] R. Sirfling. *Approximation theorems of Mathematical Statistics*. Wiley, 1980.



---

Unité de recherche INRIA Rhône-Alpes  
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Futurs : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399